Bayesian Quadratic Network Game Filters

Ceyhun Eksin, Pooya Molavi, Alejandro Ribeiro, and Ali Jadbabaie

Department of Electrical and Systems Engineering, University of Pennsylvania,

200 South 33rd Street, Philadelphia, PA 19104.

Email:{ceksin, pooya, aribeiro, jadbabai}@seas.upenn.edu.

Abstract

A repeated network game where agents have quadratic utilities that depend on information externalities – an unknown underlying state – as well as payoff externalities – the actions of all other agents in the network – is considered. Agents play Bayesian Nash Equilibrium strategies with respect to their beliefs on the state of the world and the actions of all other nodes in the network. These beliefs are refined over subsequent stages based on the observed actions of neighboring peers. This paper introduces the Quadratic Network Game (QNG) filter that agents can run locally to update their beliefs, select corresponding optimal actions, and eventually learn a sufficient statistic of the network's state. The QNG filter is demonstrated on a Cournot market competition game and a coordination game to implement navigation of an autonomous team.

I. Introduction

Games with information and payoff externalities are common models of networked economic behavior. In, e.g., trade decisions in a stock market, the payoff that a player receives depends not only on the fundamental (unknown) price of the stock but on the buy decisions of other market participants. Thus, players must respond to both, their belief on the price of the stock and their belief on the actions of other players [2]. Similar games can also be used to model the coordination of members of an autonomous team whereby agents want to select an action that is jointly optimal but only have partial knowledge about what the action of other members of the team will be. Consequently, agents select actions that they deem optimal given what they know about the task they want to accomplish and the actions they expect other agents to take.

In both of the examples in the previous paragraph we have a network of autonomous agents intent on selecting actions that maximize local utilities that depend on an unknown state of the world – information externalities – and also the unknown actions of all other agents – payoff externalities. In a Bayesian setting – or a rational setting, to use the nomenclature common in the economics literature [3] – nodes form a belief on the actions of their peers and select an action that maximizes the expected payoff with respect to those beliefs. In turn, forming these beliefs requires that each network element make a model of how other members will respond to their local beliefs. The natural assumption is that they exhibit the same behavior, namely that they are also maximizing their expected payoffs with respect to a model of other nodes' responses. But that means the first network element needs a model of other agents' models which shall include their models of his model of their model and so on. The fixed point of this iterative chain of reasoning is a Bayesian Nash Equilibrium (BNE).

In this paper we consider repeated versions of this game in which agents observe the actions taken by neighboring agents at a given time. In observing neighboring actions agents have the opportunity to learn about the private information that neighbors are, perhaps unwillingly, revealing [2]. Acquiring this information alters agents' beliefs leading to the selection of new actions which become known at the next play prompting further reevaluation of beliefs and corresponding actions. In this context we talk of Bayesian learning because the agents' goal can be reinterpreted as the eventual learning of peers' actions so that expected payoffs coincide with actual payoffs. This paper considers Gaussian prior distributions and quadratic utilities. For this type of problem we introduce the Quadratic Network Game (QNG) filter that agents can run locally to update their beliefs, select corresponding actions that maximize expected payoffs, and eventually learn a sufficient statistic of the network's state.

The burden of computing a BNE in repeated games is, in general, overwhelming even for small sized networks [4]. This intractability has led to the study of simplified models in which agents are non-Bayesian and update their beliefs according to some heuristic rule [5]–[9]. A different simplification is obtained in models with pure information externalities where payoffs depend on the self action and an underlying state but not on the actions of others. This is reminiscent of distributed estimation [10]–[19] since agents deduce the state of the world by observing neighboring actions without strategic considerations on the actions of peers. Computations are still intractable in the case of pure information externalities and for the most part only asymptotic analyses of learning dynamics with rational agents are possible [20]–[22]. Explicit methods to maximize expected payoffs given all past observations of neighboring actions are available only when signals are Gaussian [4] or when the network structure is a tree [23]. For the network games considered here in which there are information as well as payoff externalities, not much is known

besides asymptotic analyses of learning dynamics [24]–[26].

The specific setting considered in this paper is introduced in Section II. Agents repeatedly play a game whose payoffs are represented by a utility function that is quadratic in the actions of all agents and an unknown real-valued parameter. At the start of the game each agent makes a private observation of the unknown parameter corrupted by additive Gaussian noise. At each play stage agents observe actions of adjacent peers from the previous stage that they incorporate into a local observation history which they use to update their inference of the unknown parameter, and synchronously take actions that maximize their expected payoffs. Actions that maximize expected payoffs with respect to local observations histories are defined as best responses to the expected actions taken by other agents. When the expected actions of other agents are also modeled as best responses with respect to their respective observation histories, we say that the network settles into a BNE (Section II-A).

In Section III we determine a mechanism to calculate BNE actions from the perspective of an outside clairvoyant observer that knows all private observations. For this clairvoyant observer the trajectory of the game is completely determined but individual agents operate by forming a belief on the private signals of other agents. We start from the assumption that this probability distribution is normal with an expectation that, from the perspective of the outside observer, can be written as a linear combination of the actual private signals. If such is the case, we prove that there exists a set of linear equations that can be solved to obtain actions that are linear combinations of estimates of private signals (Lemma 1). This is then used to show that after observing the actions of their respective adjacent peers the probability distributions on private signals of all agents remain Gaussian with expectations that are still linear combinations of the actual private signals (Lemma 2). We proceed to close a complete induction loop to derive a recursive expression that the outside clairvoyant observer can use to compute BNE actions for all game stages (Theorem 1).

In Section IV we leverage the recursion derived in Section III to derive the QNG filter that agents can run locally, i.e., without access to all private signals, to compute their BNE action. Results in sections III and IV are generalized to the case of vector states and observations (Section V). We apply the scalar QNG filter to a Cournot competition model (Section VI) and to the coordinated movement of a team of mobile agents (Section VII).

Notation. Vectors $\mathbf{v} \in \mathbb{R}^n$ are written in boldface and matrices $A \in \mathbb{R}^{n \times m}$ in uppercase. We use $\mathbf{0}$ to denote all-zero matrices or vectors of proper dimension. If the dimension is not clear from context, we specify $\mathbf{0}_{n \times m}$. We use $\mathbf{1}$ to denote all-one matrices or vectors of proper dimension and $\mathbf{1}_{n \times m}$ to clarify

dimensions. We use $\mathbf{e}_i$ to denote the ith element of the standard orthonormal basis of $\mathbb{R}^n$ and $\bar{\mathbf{e}}_i := \mathbf{1} - \mathbf{e}_i$ to write an all-one vector with the ith component nulled.

II. Gaussian Quadratic Games

We consider games with incomplete information in which N identical agents in a network repeatedly choose actions and receive payoffs that depend on their own actions, an unknown scalar parameter $\theta \in \mathbb{R}$, and actions of all other agents. The network is represented by an undirected connected graph $G = (V, E)$ with node set $V = 1, \ldots, N$ and edge set E. The network structure restricts the information available to agent i who is assumed to observe actions of agents j in its neighborhood $n(i) := \{j : \{j, i\} \in E\}$ composed of agents that share an edge with him. The degree of node i is given by the cardinality of the set $n(i)$ and denoted as $d(i) := \# n(i)$. The neighbors of i are denoted $j_{i,1} <, \ldots, < j_{i,d(i)}$. We assume the network graph G is known to all agents.

At time $t = 0$ agent i observes a private signal $x_i \in \mathbb{R}$ which we model as being given by the unknown parameter θ contaminated with zero mean additive Gaussian noise ϵ_i,

$$x_i = \theta + \epsilon_i. \tag{1}$$

The noise variances are denoted as $c_i := \mathbf{E}\left[\epsilon_1^2\right]$ and grouped in the vector $\mathbf{c} := [c_1, \ldots, c_N]^T$ which is assumed known to all agents. The noise terms ϵ_i are further assumed independent across agents. For future reference define the vector of private signals $\mathbf{x} := [x_1, \ldots, x_N]^T \in \mathbb{R}^{N \times 1}$ grouping all local observations.

Consider a discrete time variable $t = 0, 1, 2, \ldots$ to index subsequent stages of the game. At each stage t agent i takes scalar action $a_i(t) \in \mathbb{R}$. The selection of agent i, along with the concurrent selections $a_j(t)$ of all other agents $j \in V \setminus \{i\}$ results in a payoff $u_i(a_i(t), \{a_j(t)\}_{j \in V \setminus i}, \theta)$ that agent i wants to make as large as possible. In this paper we restrict attention to quadratic payoffs which for simplicity we assume to be time invariant. Specifically, selection of actions $\{a_i = a_i(t)\}_{i \in V}$ when the state of the world is θ results in agent i experiencing a reward

$$u_i(a_i, \{a_j\}_{j \in V \setminus i}, \theta) := -\frac{1}{2} a_i^2 + \sum_{j \in V \setminus i} \beta_{ij} a_i a_j + \delta a_i \theta, \tag{2}$$

where $\beta_{ij} \in \mathbb{R}$ for all $i \in V$, $j \in V \setminus i$ and $\delta \in \mathbb{R}$ are real valued constants. Notice that since $\partial^2 u_i / \partial a_i^2 = -1 < 0$, the payoff function in (2) is strictly concave with respect to the self action a_i of agent i.

Although the goal of agent i is to select the action $a_i(t)$ that maximizes the payoff in (2), this is not possible because neither the state θ nor the actions $\{a_j(t)\}_{j \in V \setminus i}$ are known to him. Rather, agent i needs to reason about state θ and actions $\{a_j(t)\}_{j \in V \setminus i}$ based on its available information. At time $t = 0$ only the private signal x_i is known. Define then the initial information as $h_{i,0} = \{x_i\}$. The information $h_{i,0}$ is used to reason about θ and the initial actions $\{a_j(0)\}_{j \in V \setminus i}$ that other agents are to take in the initial stage of the game. At the playing of this stage, agent i observes the actions $\mathbf{a}_{n(i)}(0) := [a_{j_{i,1}}(0), \ldots, a_{j_{i,d(i)}}(0)]^T \in \mathbb{R}^{d(i) \times 1}$ of all agents in his neighborhood. These observed neighboring actions become part of the observation history $h_{i,1} = \{x_i, \mathbf{a}_{n(i)}(0)\} = \{h_{0,i}, \mathbf{a}_{n(i)}(0)\}$ which allows agent i to improve on his estimate of θ and the actions $\{a_j(1)\}_{j \in V \setminus i}$ that other agents will play on the first stage of the game, thereby also affecting the selection of its own action $a_i(1)$. In general, at any point in time t the history of observations $h_{i,t}$ is augmented to incorporate the actions of neighbors in the previous stage,

$$h_{i,t} := \{h_{i,t-1}, \mathbf{a}_{n(i)}(t-1)\} = \{x_i, \mathbf{a}_{n(i)}(u), u < t\}. \tag{3}$$

The observed action history $h_{i,t}$ is then used to update the estimates of the world state θ and the upcoming actions $\{a_j(t)\}_{j \in V \setminus i}$ of all other agents leading to the selection of the action $a_i(t)$ in the current stage of the game.

The final components of the game that we introduce are the strategies $\sigma_{i,t}$ that are used to map histories to actions. In this paper we focus on pure strategies that can be written as functions that map history realizations $h_{i,t}$ to actions $a_i(t)$

$$\sigma_{i,t} : h_{i,t} \mapsto a_i(t). \tag{4}$$

We emphasize the difference between strategy and action. An action $a_i(t)$ is the play of agent i at time t, whereas strategies $\sigma_{i,t}$ refer to the map of histories to actions. We can think of the action $a_i(t) = \sigma_{i,t}(h_{i,t})$ as the value of the strategy function $\sigma_{i,t}$ associated with the given observed history $h_{i,t}$. Further define the strategy of agent i as the concatenation $\sigma_i := \{\sigma_{i,u}\}_{u=0,\ldots,\infty}$ of strategies that agent i plays at all times. Use $\sigma_t := \{\sigma_{i,t}\}_{i \in V}$ to refer to the strategies of all players at time t, $\sigma_{0:t} := \{\sigma_u\}_{u=0,\ldots t}$ to represent the strategies played by all players between times 0 and t, and $\sigma := \{\sigma_u\}_{u=0,\ldots,\infty} = \{\sigma_i\}_{i \in V}$ to denote the strategy profile for all agents $i \in V$ and times t. As in the case of the network topology, the strategy σ is also assumed to be known to all agents. We study mechanisms for the construction of strategies in the following section.

A. Bayesian Nash equilibria

Given that agent i wants to maximize the utility in (2) but has access to the partial information available in the observed history $h_{i,t}$ in (3), a reasonable strategy $\sigma_{i,t}$ is to select the action $a_i(t)$ that maximizes the expected utility with respect to the history $h_{i,t}$. To write this formally note that this expected utility depends on strategies $\sigma_{0:t-1}$ played in the past by all agents and on strategies $\{\sigma_{j,t}\}_{j \in V \setminus i}$ that all other agents are to play in the upcoming turn. Fix then the past strategies $\sigma_{0:t-1}$ and the upcoming strategies $\{\sigma_{j,t}\}_{j \in V \setminus i}$ of other players and define the corresponding best response of player i at time t as

$$\mathrm{BR}_{i,t}\Big(\sigma_{0:t-1}, \ \{\sigma_{j,t}\}_{j \in V \setminus i}\Big) := \operatorname*{argmax}_{a_i \in \mathbb{R}} \ \mathbf{E}_{\sigma_{0:t-1}}\Big[u_i(a_i, \{\sigma_{j,t}(h_{j,t})\}_{j \in V \setminus i}, \theta)\,\big|\, h_{i,t}\Big]. \tag{5}$$

The strategies $\sigma_{0:t-1}$ in (5) played at previous times mapped respective histories $\{h_{j,u}\}_{j \in V}$ to actions $\{a_j(u)\}_{j \in V}$ for $u < t$. Therefore, the past strategies $\sigma_{0:t-1}$ determine the manner in which agent i updates his beliefs on the state of the world θ and on the histories $\{h_{j,t}\}_{j \in V \setminus i}$ observed by other agents. As per (4) the strategy profiles $\{\sigma_j(t)\}_{j \in V \setminus i}$ of other players in the current stage permit transformation of history beliefs $\{h_{j,t}\}_{j \in V \setminus i}$ into a probability distribution over respective upcoming actions $\{a_j(t)\}_{j \in V \setminus i}$. The resulting joint distribution on $\{a_j(t)\}_{j \in V \setminus i}$ and θ permits evaluation and maximization of the expectation in (5).

One can think of the profiles $\{\sigma_j(t)\}_{j \in V \setminus i}$ played by other agents in the upcoming stage as the model agent i makes of the behavior of other agents. In that sense the sensible assumption is that other agents are also playing best response to a best response model of other agents. I.e., agent i assumes agent j is playing the best response to its respective model of the behavior of other agents and that the model agent j makes of these responses is that these agents also play best response to a best response model. This modeling assumption leads to the definition of Bayesian Nash equilibria (BNE) as the solution to the fixed point equation

$$\sigma_{i,t}^*(h_{i,t}) = \mathrm{BR}_{i,t}(\sigma_{0:t-1}^*, \ \{\sigma_{j,t}^*\}_{j \in V \setminus i}), \quad \text{for all } h_{i,t}, \tag{6}$$

where we have also added the restriction that an equilibrium strategy $\sigma_{i,t-1}^*$ has been played for all times $u < t$. We emphasize that (6) needs to be satisfied for all possible histories $h_{i,t}$ and not just for the history realized in a particular game realization. This is necessary because agent i doesn't know the history observed by agent j but rather a probability distribution on histories. Thus, to evaluate the expectation in (5) agent i needs a representation of the equilibrium strategy for all possible histories $h_{j,t}$.

If all agents play their BNE strategies as defined in (6), $\sigma_{i,t}^*$ becomes optimal in the usual game

theoretic sense. There is no strategy that agent i could unilaterally deviate to that provides a higher expected payoff than $\sigma_{i,t}^*$ [cf. (5)]. In that sense the BNE strategy is the best that agent i can do given other agents' strategies and his locally available information $h_{i,t}$. In the rest of the paper we consider agents playing with respect to the BNE strategy $\sigma_{i,t}^*$ at all times. To simplify future notation define the expectation operator

$$\mathbf{E}_{i,t}\big[\ \cdot\ \big] := \mathbf{E}_{\sigma_{0:t-1}^*}\big[\ \cdot\ \mid h_{i,t}\big], \tag{7}$$

to represent expectations with respect to the local history $h_{i,t}$ when agents have played the equilibrium strategy $\sigma_{0:t-1}^*$ in all earlier stages of the game. Similarly, we define the conditional probability distribution of agent i at time t given past strategies $\sigma_{0:t-1}^*$ and his information $h_{i,t}$ as $\mathbf{P}_{i,t}(\cdot) := \mathbf{P}_{\sigma_{0:t-1}^*}\big(\cdot \mid h_{i,t}\big)$.

Since $u_i(a_i, \{a_j\}_{j\in V\setminus i}, \theta)$ is a strictly concave quadratic function of a_i as per (2), the same is true of the expected utility $\mathbf{E}_{i,t}\big[u_i(a_i, \{\sigma_{j,t}\}_{j\in V\setminus i}, \theta)\big]$ that we maximize to obtain the best response in (5). We can then rewrite (5) by nulling the derivative of the expected utility with respect to a_i. It follows that the fixed point equation in (6) can be rewritten as the set of equations

$$\sigma_{i,t}^*(h_{i,t}) = \sum_{j\in V\setminus\{i\}} \beta_{ij}\mathbf{E}_{i,t}[\sigma_{j,t}^*(h_{j,t})] + \delta\,\mathbf{E}_{i,t}[\theta], \tag{8}$$

that need to be satisfied for all possible histories $h_{i,t}$ and agents i. Our goal is to develop a filter that agents can use to compute their equilibrium actions $a_i^*(t) := \sigma_{i,t}^*(h_{i,t})$ given their observed history $h_{i,t}$. We pursue this in the following section after some remarks.

Remark 1 It may be of interest to modify the utility in (2) to include more additive terms that are functions of other actions $\{\mathbf{a}_j\}_{j\in V\setminus i}$ and the state of the world θ but not of the self actions a_i. This may change the utility and the expected utility in (5) but doesn't change the equilibrium strategy in (6). Since these terms do not contain the self action a_i, their derivatives are null and do not alter the fixed point equation in (8).

Remark 2 The equilibrium notion in (6) is based on the premise of myopic agents that choose actions that optimize payoffs at the present game stage. A more general model is to consider non-myopic agents that consider discounted payoffs of future stages. Non-myopic behavior introduces another layer of strategic reasoning. Forward looking agents would need to take into account the effect of their decisions at each stage of the game on the future path of play knowing that other agents base their future decisions on what they have previously observed. E.g., non-myopic agents might reduce their immediate payoff to harvest

information that may result in future gains. Extensions to games with non-myopic agents is beyond the scope of this paper.

III. PROPAGATION OF PROBABILITY DISTRIBUTIONS

According to the model in (8), at each stage of the game agents use the observed history $h_{i,t}$ to estimate the unknown parameter θ as well as the histories $\{h_{j,t}\}_{j \in V \setminus i}$ observed by other agents. They use the latter and the known BNE strategy $\{\sigma^*_{j,t}(h_{j,t})\}_{j \in V \setminus i}$ to form a belief $\mathbf{P}_{i,t}(\{a^*_j(t)\}_{j \in V \setminus i})$ on the actions $\{a^*_j(t)\}_{j \in V \setminus i}$ of other agents which they use to compute their equilibrium action $a^*_j(t)$ at time t. Observe that if the vector of private signals $\mathbf{x} := [x_1, \ldots, x_N]^T$ is given – not to the agents but to an outside observer – the trajectory of the game is completely determined as there are no random decisions. Thus, agent i can form beliefs on the histories $\{h_{j,t}\}_{j \in V \setminus i}$ and actions $\{a^*_j(t)\}_{j \in V \setminus i}$ of other agents if it keeps a local belief $\mathbf{P}_{i,t}(\mathbf{x})$ on the vector of private signals $\mathbf{x}$. A method to track this probability distribution is derived in this section using a complete induction argument.

Start by assuming that at given time t, the posterior distribution $\mathbf{P}_{i,t}(\mathbf{x})$ is normal. Recalling the definition of the expectation operator $\mathbf{E}_{i,t}[\,\cdot\,]$ in (7), the mean of this normal distribution is $\mathbf{E}_{i,t}[\mathbf{x}]$. Define the corresponding error covariance matrix $M^i_{\mathbf{xx}}(t) \in \mathbb{R}^{N \times N}$ as

$$M^i_{\mathbf{xx}}(t) := \mathbf{E}_{i,t}\Big[\big(\mathbf{x} - \mathbf{E}_{i,t}[\mathbf{x}]\big)\big(\mathbf{x} - \mathbf{E}_{i,t}[\mathbf{x}]\big)^T\Big]. \tag{9}$$

Although agent i's probability distribution for $\mathbf{x}$ is sufficient to describe its belief on the state of the system, subsequent derivations are simpler if we keep an explicit belief on the state of the world θ. Therefore, we also assume that agent i's beliefs on θ and $\mathbf{x}$ are jointly Gaussian given history $h_{i,t}$. The mean of θ is $\mathbf{E}_{i,t}[\theta]$ and the corresponding variance is

$$M^i_{\theta\theta}(t) := \mathbf{E}_{i,t}\Big[\big(\theta - \mathbf{E}_{i,t}[\theta]\big)\big(\theta - \mathbf{E}_{i,t}[\theta]\big)^T\Big]. \tag{10}$$

The cross covariance $M^i_{\theta\mathbf{x}}(t) \in \mathbb{R}^{1 \times N}$ between the world state θ and the private signals $\mathbf{x}$ is

$$M^i_{\theta\mathbf{x}}(t) := \mathbf{E}_{i,t}\Big[\big(\theta - \mathbf{E}_{i,t}[\theta]\big)\big(\mathbf{x} - \mathbf{E}_{i,t}[\mathbf{x}]\big)^T\Big]. \tag{11}$$

We further make the stronger assumption that the means of this joint Gaussian distribution can be written as linear combinations of the private signals. In particular, we assume that for some known matrix $L_{i,t} \in \mathbb{R}^{N \times N}$ and vector $\mathbf{k}_{i,t} \in \mathbb{R}^{N \times 1}$ we can write

$$\mathbf{E}_{i,t}[\mathbf{x}] = L_{i,t}\mathbf{x}, \qquad \mathbf{E}_{i,t}[\theta] = \mathbf{k}_{i,t}^T\mathbf{x}. \tag{12}$$

Observe that the assumption in (12) is not that the estimates $\mathbf{E}_{i,t}[\mathbf{x}]$ and $\mathbf{E}_{i,t}[\theta]$ are computed as linear combinations of the private signals $\mathbf{x}$ – indeed, $\mathbf{x}$ is not known by agent i in general. The assumption is that from the perspective of an external observer the actual computations that agents do are equivalent to the linear transformations in (12).

Under the complete induction hypothesis of Gaussian posterior beliefs at time t with expectations as in (12), we show that agents play according to linear equilibrium strategies of the form

$$\sigma_{i,t}^*(h_{i,t}) = \mathbf{v}_{i,t}^T \mathbf{E}_{i,t}[\mathbf{x}], \tag{13}$$

for some action coefficients $\mathbf{v}_{i,t} \in \mathbb{R}^{N \times 1}$ that vary across agents but are independent of the observed history $h_{i,t}$. These can be found by solving a system of linear equations. We do this in the following lemma.

Lemma 1 *Consider a Bayesian game with quadratic utility as in* (2). *Suppose that for all agents i, the joint posterior beliefs $\mathbf{P}_{i,t}([\theta, \mathbf{x}^T])$ on the state of the world θ and the private signals $\mathbf{x}$ given the local history $h_{i,t}$ at time t are Gaussian with means expressed as the linear combinations of private signals in* (12) *for some known vectors $\mathbf{k}_{i,t}$ and matrices $L_{i,t}$. Define the aggregate vector $\mathbf{k}_t := [\mathbf{k}_{1,t}^T, \ldots, \mathbf{k}_{N,t}^T]^T \in \mathbb{R}^{N^2 \times 1}$ stacking the state estimation weights of all agents and the block matrix $L_t \in \mathbb{R}^{N^2 \times N^2}$ with $N \times N$ diagonal blocks $((L_t))_{ii} = L_{i,t}^T$ and off diagonal blocks $((L_t))_{ij} = -\beta_{ij} L_{i,t}^T L_{j,t}^T$,*

$$L_t := \begin{pmatrix} L_{1,t}^T & -\beta_{12} L_{1,t}^T L_{2,t}^T & \ldots & -\beta_{1N} L_{1,t}^T L_{N,t}^T \\ -\beta_{21} L_{2,t}^T L_{1,t}^T & L_{2,t}^T & \ldots & -\beta_{2N} L_{2,t}^T L_{N,t}^T \\ \vdots & \ldots & \ddots & \vdots \\ -\beta_{N-11} L_{N-1,t}^T L_{1,t}^T & \ldots & L_{N-1,t}^T & -\beta_{N-1N} L_{N-1,t}^T L_{N,t}^T \\ -\beta_{N1} L_{N,t}^T L_{1,t}^T & \ldots & -\beta_{NN-1} L_{N,t}^T L_{N-1,t}^T & L_{N,t}^T \end{pmatrix}. \tag{14}$$

If there exists a linear equilibrium strategy as in (13), *the action coefficients $\mathbf{v}_t := [\mathbf{v}_{1,t}^T, \ldots, \mathbf{v}_{N,t}^T]^T \in \mathbb{R}^{N^2}$ can be obtained by solving the system of linear equations*

$$L_t \mathbf{v}_t = \delta \mathbf{k}_t. \tag{15}$$

Proof: We hypothesize that agents play according to a linear equilibrium strategy as in (13). Substituting this candidate strategy into the equilibrium equations in (8) yields

$$\mathbf{v}_{i,t}^T \mathbf{E}_{i,t}[\mathbf{x}] = \sum_{j \in V \setminus \{i\}} \beta_{ij} \mathbf{E}_{i,t} \left[\mathbf{v}_{j,t}^T \mathbf{E}_{j,t}[\mathbf{x}] \right] + \delta \mathbf{E}_{i,t}[\theta]. \tag{16}$$

The summation in (16) includes the expectations $\mathbf{E}_{i,t}[\mathbf{E}_{j,t}[\mathbf{x}]]$ of agent i on the private signals' estimate of agent j. As per the induction hypothesis in (12), we have that the inner expectations can be written

as $\mathbf{E}_{j,t}[\mathbf{x}] = L_{j,t}\mathbf{x}$. Using this fact, agent i's expectation of agent j's estimate of private signals becomes

$$\mathbf{E}_{i,t}\Big[\mathbf{E}_{j,t}[\mathbf{x}]\Big] = L_{j,t}\mathbf{E}_{i,t}[\mathbf{x}]. \tag{17}$$

Substituting (17) and the estimate induction hypotheses in (12) for the corresponding terms in (16) and (17), and reordering terms yield the set of equations

$$\mathbf{v}_{i,t}^T L_{i,t}\mathbf{x} = \sum_{j \in V \setminus \{i\}} \beta_{ij}\mathbf{v}_{j,t}^T L_{j,t}L_{i,t}\mathbf{x} + \delta\,\mathbf{k}_{i,t}^T\mathbf{x}, \tag{18}$$

At this point we recall that the equilibrium equations in (8) are true for all possible histories $h_{i,t}$. Therefore, the equilibrium equations in (18), which are derived from (8), have to hold irrespectively of the history's realization. This in turn means that they will be true for all possible values of $\mathbf{x}$. This can be ensured by equating the coefficients that multiply each component of $\mathbf{x}$ in (18) thereby yielding the relationships

$$L_{i,t}^T\mathbf{v}_{i,t} = \sum_{j \in V \setminus \{i\}} \beta_{ij} L_{i,t}^T L_{j,t}^T\mathbf{v}_{j,t} + \delta\,\mathbf{k}_{i,t}, \tag{19}$$

that need to hold true for all agents i. The result in (15) is just a restatement of (19) with the latter corresponding to the i-th block of the relationship in (15). ∎

Lemma 1 provides a mechanism to determine the strategy profiles $\sigma_{i,t}^*(\cdot)$ of all agents through the computation of the action vectors $\mathbf{v}_{i,t}$ as a block of the vector $\mathbf{v}_t$ that solves (15). We emphasize that the value of the weight vector $\mathbf{v}_t$ in (15) does not depend on the realization of private signals $\mathbf{x}$. This is as it should be because the postulated equilibrium strategy in (13) assumes the action weights $\mathbf{v}_{i,t}$ are independent of the observed history. A consequence of this fact is that the action coefficients $\{\mathbf{v}_{i,t}\}_{i \in V}$ of all agents can be determined locally by all agents as long as the matrices $L_{i,t}$ and vectors $\mathbf{v}_{i,t}$ are common knowledge. The equilibrium actions $a_i^*(t)$, however, do depend on the observed history because to determine the action $a_i^*(t) = \sigma_{i,t}^*(h_{i,t}) = \mathbf{v}_{i,t}^T\mathbf{E}_{i,t}[\mathbf{x}]$ we multiply $\mathbf{v}_{i,t}^T$ by the expectation $\mathbf{E}_{i,t}[\mathbf{x}]$ associated with the actual observed history $h_{i,t}$. See Section IV for details.

At time t agent i computes its action vector $\mathbf{v}_{i,t}$ which it uses to select the equilibrium action $a_i^*(t) = \mathbf{v}_{i,t}^T\mathbf{E}_{i,t}[\mathbf{x}]$ as per (13). Since we have also hypothesized that $\mathbf{E}_{i,t}[\mathbf{x}] = L_{i,t}\mathbf{x}$, as per (12) the action of agent i at time t is given by

$$a_i(t) = \mathbf{v}_{i,t}^T L_{i,t}\mathbf{x}. \tag{20}$$

We emphasize that as in (12) the expression in (20) is not the computation made by agent i but an equivalent computation from the perspective of an external omniscient observer.

The actions $\mathbf{a}_{n(i)}(t) := [a_{j_{i,1}}(t), \ldots, a_{j_{i,d(i)}}(t)]^T \in \mathbb{R}^{d(i) \times 1}$ of neighboring agents $j \in n(i)$ become part of the observed history $h_{i,t+1}$ of agent i at time $t+1$ [cf. (3)]. The important consequence of (20) is that these observations are a linear combination of private signals $\mathbf{x}$. In particular, by defining the matrix $H_{i,t}^T := [\mathbf{v}_{j_{i,1},t}^T L_{j_{i,1},t}; \ldots; \mathbf{v}_{j_{i,d(i)},t}^T L_{j_{i,d(i)},t}] \in \mathbb{R}^{d(i) \times N}$ we can write

$$\mathbf{a}_{n(i)}(t) = H_{i,t}^T \mathbf{x} := \begin{pmatrix} \mathbf{v}_{j_{i,1},t}^T L_{j_{i,1},t} \\ \vdots \\ \mathbf{v}_{j_{i,d(i)},t}^T L_{j_{i,d(i)},t} \end{pmatrix} \mathbf{x}. \tag{21}$$

Agent i's belief of $\mathbf{x}$ at time t is normally distributed; moreover, when we go from time t to time $t+1$, agent i observes a linear combination, $\mathbf{a}_{n(i)}(t) = H_{i,t}^T \mathbf{x}$, of private signals. Thus, the propagation of the probability distribution when the history $h_{i,t+1}$ incorporates the actions $\mathbf{a}_{n(i)}(t)$ is a simple sequential LMMSE estimation problem [27, Ch. 12]. In particular, the joint posterior distribution of $\mathbf{x}$ and θ given $h_{i,t+1}$ remains Gaussian and the expectations $\mathbf{E}_{i,t+1}[\mathbf{x}]$ and $\mathbf{E}_{i,t+1}[\theta]$ remain linear combinations of private signals $\mathbf{x}$ as in (12) for some matrix $L_{i,t+1}$ and vector $\mathbf{k}_{i,t+1}$ which we compute explicitly in the following lemma.

Lemma 2 *Consider a Bayesian game with quadratic utility as in* (2) *and the same assumptions and definitions of Lemma 1. Further define the observation matrix* $H_{i,t}^T := [\mathbf{v}_{j_{i,1},t}^T L_{j_{i,1},t}; \ldots; \mathbf{v}_{j_{i,d(i)},t}^T L_{j_{i,d(i)},t}] \in \mathbb{R}^{d(i) \times N}$ *as in* (21) *and the LMMSE gains*

$$K_{\mathbf{x}}^i(t) := M_{\mathbf{xx}}^i(t) H_{i,t} \left(H_{i,t}^T M_{\mathbf{xx}}^i(t) H_{i,t} \right)^{-1}, \tag{22}$$

$$K_{\theta}^i(t) := M_{\theta \mathbf{x}}^i(t) H_{i,t} \left(H_{i,t}^T M_{\mathbf{xx}}^i(t) H_{i,t} \right)^{-1}, \tag{23}$$

and assume that agents play the linear equilibrium strategy in (13). *Then, the beliefs* $\mathbf{P}_{i,t+1}([\theta, \mathbf{x}^T])$ *after observing neighboring actions at time* t *are Gaussian with means that can be expressed as the linear combination of private signals*

$$\mathbf{E}_{i,t+1}[\mathbf{x}] = L_{i,t+1} \mathbf{x}, \qquad \mathbf{E}_{i,t+1}[\theta] = \mathbf{k}_{i,t+1}^T \mathbf{x}, \tag{24}$$

where the matrix $L_{i,t+1}$ *and vector* $\mathbf{k}_{i,t+1}$ *are given by*

$$L_{i,t+1} = L_{i,t} + K_{\mathbf{x}}^i(t) \left(H_{i,t}^T - H_{i,t}^T L_{i,t} \right), \tag{25}$$

$$\mathbf{k}_{i,t+1}^T = \mathbf{k}_{i,t}^T + K_{\theta}^i(t) \left(H_{i,t}^T - H_{i,t}^T L_{i,t} \right). \tag{26}$$

The posterior covariance matrix $M_{\mathbf{xx}}^i(t+1)$ for the private signals $\mathbf{x}$ the variance $M_{\theta\theta}^i(t+1)$ of the state θ and the cross covariance $M_{\theta\mathbf{x}}^i(t+1)$ are further given by

$$M_{\mathbf{xx}}^i(t+1) = M_{\mathbf{xx}}^i(t) - K_{\mathbf{x}}^i(t)H_{i,t}^T M_{\mathbf{xx}}^i(t), \tag{27}$$

$$M_{\theta\theta}^i(t+1) = M_{\theta\theta}^i(t) - K_{\theta}^i(t)^T H_{i,t}^T M_{\mathbf{x}\theta}^i(t), \tag{28}$$

$$M_{\theta\mathbf{x}}^i(t+1) = M_{\theta\mathbf{x}}^i(t) - K_{\theta}^i(t)H_{i,t}^T M_{\mathbf{xx}}^i(t). \tag{29}$$

Proof: Since observations of i, $a_{n(i)}(t)$, are linear combinations of private signals $\mathbf{x}$ which are normally distributed, observations of i are also normally distributed from the perspective of i. Furthermore, by assumption (12), the prior distribution $\mathbf{P}_{i,t}(\mathbf{x})$ is Gaussian. Hence, the posterior distribution, $\mathbf{P}_{i,t+1}(\mathbf{x})$, is also Gaussian. Specifically, the mean of the posterior distribution corresponds to the LMMSE estimator with gain matrix $K_{\mathbf{x}}^i(t) = M_{\mathbf{xx}}^i(t)H_{i,t}\big(H_{i,t}^T M_{\mathbf{xx}}^i(t)H_{i,t}\big)^{-1}$; that is,

$$\mathbf{E}_{i,t+1}[\mathbf{x}] = \mathbf{E}_{i,t}[\mathbf{x}] + K_{\mathbf{x}}^i(t)\big(\mathbf{a}_{n(i)}(t) - \mathbf{E}_{i,t}[\mathbf{a}_{n(i)}(t)]\big). \tag{30}$$

Because θ and $\mathbf{x}$ are jointly Gaussian at time t, θ and $\mathbf{a}_{n(i)}(t)$ are also jointly Gaussian. Therefore, the posterior distribution $\mathbf{P}_{i,t+1}(\theta)$ is also Gaussian. Consequently, the Bayesian estimate of θ is given by a sequential LMMSE estimator with gain matrix $K_{\theta}^i(t) = M_{\theta\mathbf{x}}^i(t)H_{i,t}\big(H_{i,t}^T M_{\mathbf{xx}}^i(t)H_{i,t}\big)^{-1}$,

$$\mathbf{E}_{i,t+1}[\theta] = \mathbf{E}_{i,t}[\theta] + K_{\theta}^i(t)\big(\mathbf{a}_{n(i)}(t) - \mathbf{E}_{i,t}[\mathbf{a}_{n(i)}(t)]\big). \tag{31}$$

Given the linear observation model in (21), agent i's estimate of his observations at time t is given by $\mathbf{E}_{i,t}(\mathbf{a}_{n(i)}(t)) = H_{i,t}^T \mathbf{E}_{i,t}[\mathbf{x}]$. Substituting (12) for the mean estimates at time t in (30) and (31), we obtain

$$\mathbf{E}_{i,t+1}[\mathbf{x}] = L_{i,t}\mathbf{x} + K_{\mathbf{x}}^i(t)\big(H_{i,t}^T\mathbf{x} - H_{i,t}^T L_{i,t}\mathbf{x}\big), \tag{32}$$

$$\mathbf{E}_{i,t+1}[\theta] = \mathbf{k}_{i,t}^T\mathbf{x} + K_{\theta}^i(t)\big(H_{i,t}^T\mathbf{x} - H_{i,t}^T L_{i,t}\mathbf{x}\big). \tag{33}$$

Grouping the terms that multiply $\mathbf{x}$ on the right hand side of the two equations, we observe that $\mathbf{E}_{i,t+1}[\mathbf{x}] = L_{i,t+1}\mathbf{x}$ and $\mathbf{E}_{i,t+1}[\theta] = \mathbf{k}_{i,t+1}^T\mathbf{x}$ where $L_{i,t+1}$ and $\mathbf{k}_{i,t+1}$ are as defined in (25) and (26). Similarly, the updates for error covariance matrices are as given in (27)–(29) following standard LMMSE updates [27, Ch. 12]. ∎

In the repeated game we are considering, agents determine optimal actions given available information and determine the information that is revealed by neighboring actions. These questions are respectively answered by lemmas 1 and 2 under the inductive hypotheses of Gaussian beliefs and linear estimates as

per (12). The answer provided by Lemma 2 also shows that the inductive hypotheses hold true at time $t + 1$ and provides an explicit recursion to propagate the mean and variance of the beliefs posterior to the observation of neighboring actions. This permits closing the inductive loop to establish the following theorem for recursive computation of BNE of repeated games with quadratic payoffs.

Theorem 1 *Consider a repeated Bayesian game with the quadratic utility function in (2) and assume that linear strategies $\sigma_{i,t}^*(h_{i,t}) = \mathbf{v}_{i,t}^T \mathbf{E}_{i,t}[\mathbf{x}]$ as in (13) exist for all times t. Then, the action coefficients $\mathbf{v}_{i,t}$ can be computed by solving the system of linear equations in (15) with $\mathbf{v}_t := [\mathbf{v}_{1,t}^T, \dots, \mathbf{v}_{N,t}^T]^T$, $\mathbf{k}_t := [\mathbf{k}_{1,t}^T, \dots, \mathbf{k}_{N,t}^T]^T$ and L_t as in (14). The matrices $L_{i,t}$ and the vectors $\mathbf{k}_{i,t}$ are computed by recursive application of (22)-(23) and (25)-(29) with initial values*

$$L_{i,0} = \mathbf{1}\mathbf{e}_i^T, \qquad \mathbf{k}_{i,0} = \mathbf{e}_i. \tag{34}$$

The initial covariance matrix $M_{\mathbf{xx}}^i(0)$, initial variance $M_{\theta\theta}^i(0)$, and initial cross covariance $M_{\theta\mathbf{x}}^i(0)$ are given by

$$M_{\mathbf{xx}}^i(0) = \mathrm{diag}(\bar{\mathbf{e}}_i)\mathrm{diag}(\mathbf{c}) + \bar{\mathbf{e}}_i\bar{\mathbf{e}}_i^T c_i, \quad M_{\theta\theta}^i(0) = c_i, \quad M_{\theta\mathbf{x}}^i(0) = c_i\bar{\mathbf{e}}_i^T. \tag{35}$$

Proof: See Appendix A. ∎

According to Theorem 1, the beliefs on θ and $\mathbf{x}$ remain Gaussian for all agents and all times when agents play according to a linear equilibrium strategy as in (13) at each stage. Theorem 1 also provides a recursive mechanism to compute the coefficients $\mathbf{v}_{i,t}$ of the linear BNE strategies $\sigma_{i,t}^*(h_{i,t}) = \mathbf{v}_{i,t}^T \mathbf{E}_{i,t}[\mathbf{x}]$ and the coefficients $L_{i,t}$ and $\mathbf{k}_{i,t}$ that determine the LMMSE estimates as per (12). However, these latter expressions cannot be used by agent i to calculate estimates $\mathbf{E}_{i,t}[\mathbf{x}]$ and $\mathbf{E}_{i,t}[\theta]$ unless the private signals $\mathbf{x}$ are exactly known, which will absolve agent i from responsibility of the estimation process entirely. Since the BNE action $a_i^*(t) = \sigma_{i,t}^*(h_{i,t}) = \mathbf{v}_{i,t}^T \mathbf{E}_{i,t}[\mathbf{x}]$ depends on having the observed private signal estimate $\mathbf{E}_{i,t}[\mathbf{x}]$ available, Theorem 1 does not provide a way of computing the optimal action either. This mismatch can be solved by writing the LMMSE updates in a different form as we show in the next section after the following remark.

Remark 3 Results in this paper assume the system of linear equations in (15) has a unique solution. If the solution is not unique, a prior agreement is necessary for agents to play consistent strategies. E.g., agents could agree beforehand to select the vector $\mathbf{v}_t$ with minimum Euclidean norm. If (15) does not have a solution, it means that the equilibrium strategies of the form in (20) do not exist. A sufficient
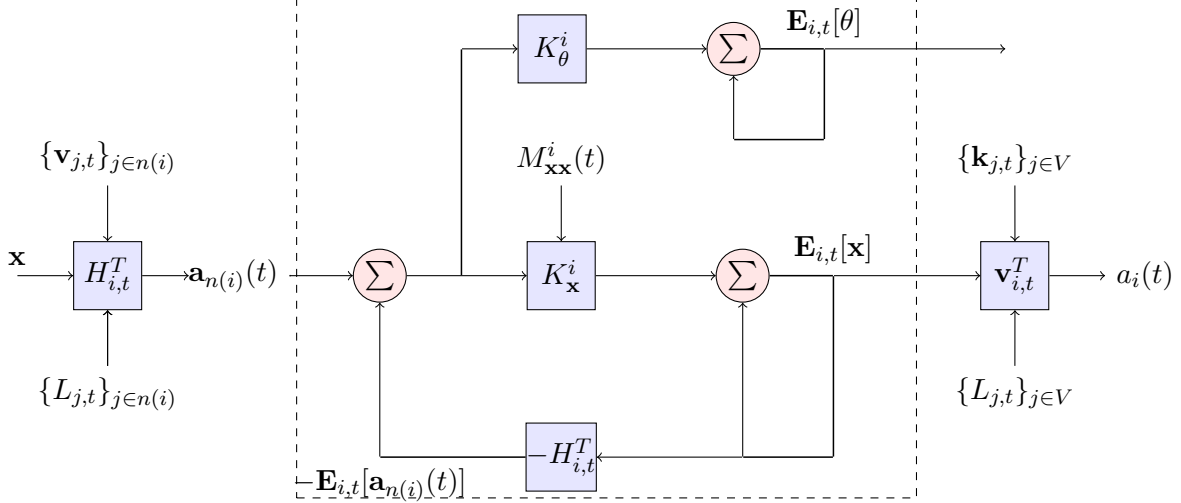
Fig. 1. Quadratic Network Game (QNG) filter at agent i. There are two types of blocks, circle and rectangle. Arrows coming into the circle block are summed. The arrow that goes into a rectangle block is multiplied by the coefficient written inside the block. Inside the dashed box agent i's mean estimate updates on $\mathbf{x}$ and θ are illustrated (cf. (36) and (37)). The gain coefficients for the mean updates are fed from LMMSE block in Fig. 2. The observation matrix $H_{i,t}$ is fed from the game block in Fig. 2. Agent i multiplies his mean estimate on $\mathbf{x}$ at time t with action coefficient $\mathbf{v}_{i,t}$, which is fed from game block in Fig. 2, to obtain $a_i(t)$. The mean estimates $\mathbf{E}_{i,t}[\mathbf{x}]$ and $a_i(t)$ can only be calculated by agent i.

condition for this *not* to happen is to have a strictly diagonally dominant utility function which in explicit terms we write $\sum_{j \in V \setminus \{i\}} |\beta_{ij}| < 1$. In this case Gershgorin's Theorem implies that L_t is full rank because it has no null eigenvalues. Laxer conditions to guarantee existence of linear equilibria as in (20) can be found in, e.g., [28], [29]. In all of our numerical experiments solutions to (15) exist and are unique.

IV. QUADRATIC NETWORK GAME FILTER

To compute and play BNE strategies each node runs the quadratic network game (QNG) filter that we derive in this section. Since agent i cannot use (12), we need an alternative means of computing estimates $\mathbf{E}_{i,t}[\mathbf{x}]$ and $\mathbf{E}_{i,t}[\theta]$. To do this refer to the transformation of (30) and (31) into (32) and (33) in the proof of Lemma 2. In this transformation we substitute the observed neighboring actions $\mathbf{a}_{n(i)}(t)$ for their model $\mathbf{a}_{n(i)}(t) = H_{i,t}^T \mathbf{x}$ and write the expectation of these actions as $H_{i,t}^T \mathbf{E}_{i,t}[\mathbf{x}]$ with the further substitution $\mathbf{E}_{i,t}[\mathbf{x}] = L_{i,t}\mathbf{x}$. As a result we can rewrite (30) and (31) as

$$\mathbf{E}_{i,t+1}[\mathbf{x}] = \mathbf{E}_{i,t}[\mathbf{x}] + K_{\mathbf{x}}^i(t)\big(\mathbf{a}_{n(i)}(t) - H_{i,t}^T \mathbf{E}_{i,t}[\mathbf{x}]\big), \tag{36}$$

$$\mathbf{E}_{i,t+1}[\theta] = \mathbf{E}_{i,t}[\theta] + K_{\theta}^i(t)\big(\mathbf{a}_{n(i)}(t) - H_{i,t}^T \mathbf{E}_{i,t}[\mathbf{x}]\big). \tag{37}$$

The updates in (36) and (37) can be implemented locally by agent i since they depend on the previous values $\mathbf{E}_{i,t}[\mathbf{x}]$ and $\mathbf{E}_{i,t}[\theta]$ of the LMMSE estimates, and the observed neighboring actions $\mathbf{a}_{n(i)}(t)$. They

can be combined with the coefficient recursions in (15), (22)-(23), and (25)-(29) as well as with the BNE strategy expression in (13) to recursively compute the equilibrium actions $a_i^*(t)$ given the observed history $h_{i,t}$.

The updates in (13), (15), (22)-(23), (25)-(29), and (36)-(37) form the QNG filter. In the QNG filter agent i performs a full network simulation in which it maintains a belief $\mathbf{P}_{i,t}([\theta, \mathbf{x}^T])$ on the state of the world θ and the private signals $\mathbf{x}$ of all agents. This implies performing the coefficient updates (15), (22)-(23), (25)-(29) for all agents in the network. This he can do because the network topology and private signal models are common knowledge. The updates (13) and (36)-(37) are performed for agent i's own index only.

The signal updates on (36)-(37) are illustrated inside the dashed box in Fig. 1. At time t, the inputs to the filter are the observed actions $\mathbf{a}_{n(i)}(t)$ of agent i's neighbors. The prediction $\mathbf{E}_{i,t}[\mathbf{a}_{n(i)}(t)] = H_{i,t}\mathbf{E}_{i,t}[\mathbf{x}]$ of this vector is subtracted from the observed value and the resultant error is fed into two parallel blocks respectively tasked with updating the belief $\mathbf{E}_{i,t}[\theta]$ on the state of the world θ, and the belief $\mathbf{E}_{i,t}[\mathbf{x}]$ on the private signals $\mathbf{x}$ of other agents. The error $\mathbf{a}_{n(i),t} - \mathbf{E}_{i,t}[\mathbf{a}_{n(i),t}]$ is multiplied by the gain $K_{\mathbf{x}}^i(t)$ and the resultant innovation is added to the previous mean estimate to correct the estimate of $\mathbf{x}$ [cf. (36)]. Similarly, the error is multiplied by the gain $K_\theta^i(t)$ and the resultant innovation is added to the previous mean estimate to correct the estimate of θ at i [cf. (37)]. In order to determine the equilibrium play as per (13), agent i multiples his private signal estimate $\mathbf{E}_{i,t}[\mathbf{x}]$ by the vector $\mathbf{v}_i(t)$ obtained by solving the system of linear equations in (15).

Observe that in the QNG filter, we do not use the fact that estimates $\mathbf{E}_{i,t}[\theta]$ and $\mathbf{E}_{i,t}[\mathbf{x}]$ as well as actions $a_{i,t}$ can be written as linear combinations of the private signals [cf. (12) and (20)]. While the expressions in (12) and (20) are certainly correct, they cannot be used for implementation because $\mathbf{x}$ is only partially unknown to agent i. The role of (12) and (20) is to allow derivation of recursions that we use to keep track of the gains used in the QNG filter. These recursions can be divided into a group of LMMSE updates and a group of game updates as we show in Fig. 2.

As it follows from (22)-(23) and (27)-(29), the update of LMMSE coefficients is identical to the gain and covariance updates of a sequential LMMSE. The only peculiarity is that the observation matrix $H_{j,t}$ is fed from the game update block and is partially determined by the LMMSE gains and covariances of previous iterations. Nevertheless, this peculiarity is more associated with the game block than with the LMMSE block. The game block uses (25) and (26) to keep track of the matrices $L_{j,t}$ and the vectors $\mathbf{k}_{j,t}$. The matrices $L_{j,t}$ are used as building blocks of the matrix L_t and the vectors $\mathbf{k}_{j,t}$ are stacked in the vector $\mathbf{k}_t$ and used to formulate the systems of equations in (15). Solving this system of equations,

using L_t^{-1} when it is full rank or its pseudo inverse when it is not, yields the coefficients $\mathbf{v}_{j,t}$ which in turn determine the observation matrix $H_{j,t}$ as per (21). As mentioned before, the game block feeds the matrices $H_{j,t}$ to the filter block as they are used in the LMMSE gains and covariance updates. The LMMSE block feeds the gains $K_{\mathbf{x}}^j(t)$ and $K_{\theta}^j(t)$ to the game block as these are needed to update $L_{j,t}$ and $\mathbf{k}_{j,t}$.

We remark that agent i is keeping track of the matrices and vectors in Fig. 2 for all $i \in V$. I.e., agent i calculates observation matrices $H_{j,t}$ for $j \in V$ in the game block which are fed into the LMMSE block to obtain gains matrices $K_{\mathbf{x}}^j(t)$ and $K_{\theta}^j(t)$ for all $j \in V$. These gains are fed into the game block from the LMMSE block as they are needed to update $L_{j,t}$ and $\mathbf{k}_{j,t}$ for all $j \in V$. The reason for this is the step in the game block in which we compute the play coefficients $\mathbf{v}_{j,t}$. To solve this system of equations, agent i needs to build the matrix L_t that is formed by the blocks $L_{j,t}$ of all agents. All of these computations for the coefficients of other agents are internal to agent i and independent of the game realization. The gains can be computed offline prior to running the game.

Remark 4 The QNG filter can also be used in repeated games with purely informational externalities. In this case each agent's payoff is given by $u(\theta, a_i) = -(\theta - a_i)^2$, and the problem is thus equivalent to the distributed estimation of the world state θ [4]. Our model subsumes the games with purely informational externalities as a special case. Given this payoff function, the best response of agent i at time t is the action $a_i(t) = \mathbf{E}_{i,t}[\theta]$. Hence, it is not necessary to solve (15) for the optimal strategy coefficients $\mathbf{v}_{i,t}$. Other than this the QNG filter remains unchanged. Since in the case of purely informational externalities the end goal is the estimation of θ, the QNG filter is tantamount to an optimal distributed implementation of a Kalman filter.

V. VECTOR STATES AND VECTOR OBSERVATIONS

Consider the case when state of the world is a vector, that is, $\boldsymbol{\theta} \in \mathbb{R}^m$ for $m > 1$. Similar to the scalar case, each agent receives initial private signal $\mathbf{x}_i \in \mathbb{R}^m$,

$$\mathbf{x}_i = \boldsymbol{\theta} + \boldsymbol{\epsilon}_i \tag{38}$$

where the additive noise term $\boldsymbol{\epsilon}_i \in \mathbb{R}^m$ is multivariate Gaussian with zero mean and variance-covariance matrix $C_i \in \mathbb{R}^{m \times m}$. For future reference, define the vector obtained by stacking elements at the kth row and lth column of variance-covariance matrices of all agents, $\mathbf{C}_{k,l} := [C_1[k,l], \dots, C_N[k,l]]^T$. We use $\mathbf{x}_i[n]$ to denote the nth private signal of agent i where $n \leq m$. We assume that private signals are
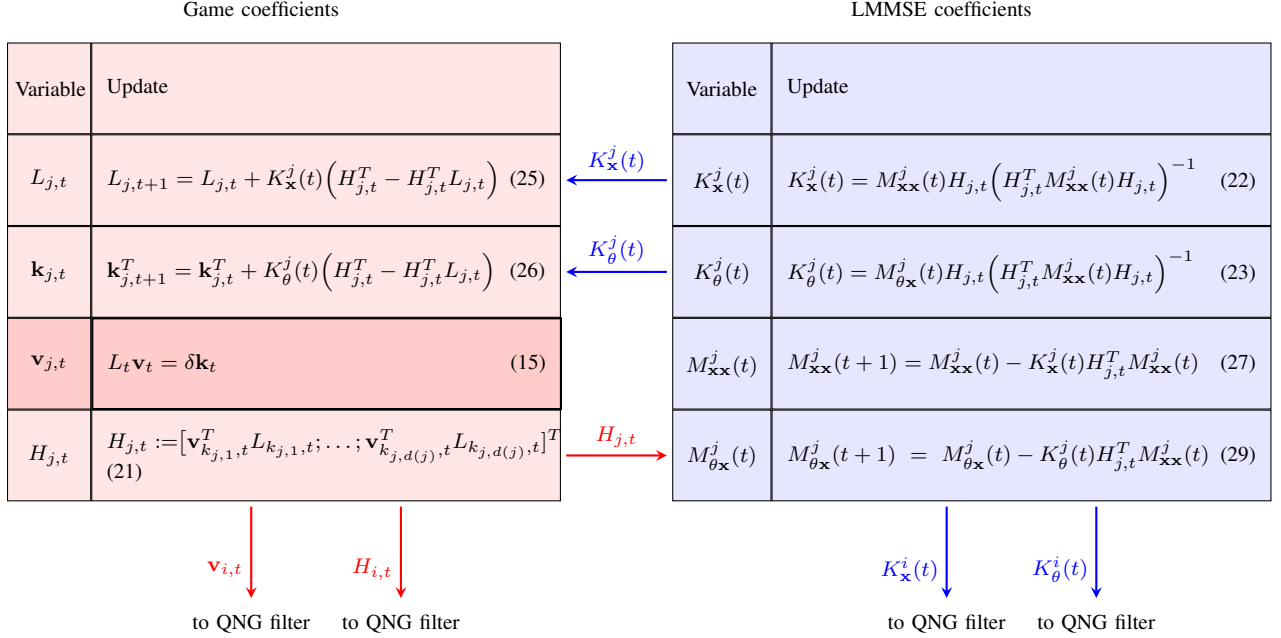
Fig. 2. Propagation of gains required to implement the Quadratic Network Game (QNG) filter of Fig. 1. Gains are separated into interacting LMMSE and game blocks. All agents perform a full network simulation in which they compute the gains of all other agents. This is necessary because when we compute the play coefficients $\mathbf{v}_{j,t}$ in the game block, agent i builds the matrix L_t that is formed by the blocks $L_{j,t}$ of all agents [cf. (14)]. This full network simulation is possible because the network topology and private signal models are common knowledge.

independent among agents, that is, $\mathbf{E}_{i,0}[\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_j] = 0$ for all $i \in V$ and $j \in V \setminus \{i\}$. We define the set of all private signals as

$$\mathbf{x} := [\mathbf{x}_1[1], \ldots, \mathbf{x}_N[1], \ldots, \mathbf{x}_1[m], \ldots, \mathbf{x}_N[m]]^T, \tag{39}$$

where $\mathbf{x} \in \mathbb{R}^{Nm \times 1}$. We use $\mathbf{x}[n] := [\mathbf{x}_1[n], \ldots, \mathbf{x}_N[n]]^T$ to denote the vector of private signals of agents on the nth state of the world.

At each stage t, agent i takes action $\mathbf{a}_i(t) \in \mathbb{R}^m$. Agent i's action at time t is to maximize a payoff function which is represented by the following quadratic function

$$u_i(\mathbf{a}_i, \{a_j\}_{j \in V \setminus i}, \boldsymbol{\theta}) = -\frac{1}{2} \sum_{j \in V} \mathbf{a}_j^T \mathbf{a}_j + \sum_{j \in V \setminus \{i\}} \mathbf{a}_i^T B_{ij} \mathbf{a}_j + \mathbf{a}_i^T D \boldsymbol{\theta}, \tag{40}$$

where constants B_{ij} and D belong to $\mathbb{R}^{m \times m}$. Similar to the scalar case, other additive terms that depend on $\{a_j\}_{j \in V \setminus i}$ and $\boldsymbol{\theta}$ can exist without changing the results to follow. We obtain the best response function for agent i by taking the derivative of the expected utility function with respect to $\mathbf{a}_i$, equating it to zero,

and solving for $\mathbf{a}_i$:

$$\text{BR}_{i,t}(\{\sigma_{j,t}(h_{j,t})\}_{j \in V \setminus i}) = \sum_{j \in V \setminus i} B_{ij} \mathbf{E}_{i,t}[\sigma_{j,t}(h_{j,t})] + D\mathbf{E}_{i,t}[\boldsymbol{\theta}]. \tag{41}$$

Note that $\text{BR}_t : \mathbb{R}^{Nm} \to \mathbb{R}^{Nm}$.

Similar to the case when the unknown parameter is a scalar, it is sufficient for agents to keep track of estimates of $\mathbf{x}$ in order to achieve the best estimate of $\boldsymbol{\theta}$. Accordingly, the definitions of estimates of private signals and the unknown parameters and their corresponding covariance matrices (9)–(11) are the same as in the scalar case.

In what follows, we show that the mean estimates are linear in private signals and equilibrium actions are linear in expectations of private signals in the similar fashion we did for the scalar state of the world.

Lemma 3 *Consider a Bayesian game with quadratic utility as in* (40). *Suppose that for all agents i, the joint posterior beliefs on the state of the world $\boldsymbol{\theta}$ and the private signals $\mathbf{x}$ given the local history $h_{i,t}$ at time t, $\mathbf{P}_{i,t}([\boldsymbol{\theta}^T, \mathbf{x}^T])$, are Gaussian with means expressed as*

$$\mathbf{E}_{i,t}[\boldsymbol{\theta}] = Q_{i,t}\mathbf{x}, \text{ and } \mathbf{E}_{i,t}[\mathbf{x}] = L_{i,t}\mathbf{x}, \tag{42}$$

where $L_{i,t} \in \mathbb{R}^{Nm \times Nm}$ and $Q_{i,t} \in \mathbb{R}^{m \times Nm}$ are known estimation weights. If there exists an equilibrium strategy profile that is linear in expectations of private signals,

$$\sigma_{i,t}^*(h_{i,t}) = U_{i,t}\mathbf{E}_{i,t}[\mathbf{x}] \quad \text{for all } i \in V, \tag{43}$$

then the action coefficients $\{U_{i,t}\}_{i \in V}$ can be obtained by solving the system of linear equations

$$L_{i,t}^T U_{i,t}^T = \sum_{j \in V \setminus i} L_{i,t}^T L_{j,t}^T U_{j,t}^T B_{ij}^T + Q_{i,t}^T D^T, \quad \text{for all } i \in V \tag{44}$$

Proof: The proof is analogous to the proof of Lemma 1. By substituting the candidate strategies in (43) to the best response function in (41) for all $i \in V$, we obtain the following equilibrium equations

$$U_{i,t}\mathbf{E}_{i,t}[\mathbf{x}] = \sum_{j \in V \setminus \{i\}} B_{ij} \mathbf{E}_{i,t}[U_{j,t}\mathbf{E}_{j,t}[\mathbf{x}]] + D\mathbf{E}_{i,t}[\boldsymbol{\theta}]. \tag{45}$$

for all $i \in V$. After using the fact that $\mathbf{E}_{i,t}[\mathbf{E}_{j,t}[\mathbf{x}]] = L_{j,t}\mathbf{E}_{i,t}[\mathbf{x}]$ with mean estimate assumptions in (42)

for the corresponding terms in (45), we obtain the following set of equations

$$U_{i,t}L_{i,t}\mathbf{x} = \sum_{j \in V \setminus \{i\}} B_{ij}U_{j,t}L_{j,t}L_{i,t}\mathbf{x} + DQ_{i,t}\mathbf{x}. \tag{46}$$

We ensure that the strategies in (43) satisfy the equilibrium equations for any realization of history by equating coefficients that multiply each component of $\mathbf{x}$ in (46) which yields the set of equations given by (44). ∎

For a linear equilibrium strategy, the actions can be written as a linear combination of the private signals using (42), that is, the action of agent i at time t is given by

$$a_i(t) = U_{i,t}L_{i,t}\mathbf{x} \qquad \text{for all } i \in V. \tag{47}$$

Being able to express actions as in (47) permits writing observations of agents in linear form. From the perspective of an observer, the action $\mathbf{a}_j(t)$ is equivalent to observing a linear combination of private signals. As a result, we can represent observation vector of agent i $\mathbf{a}_{n(i)}(t) := \left[\mathbf{a}_{j_1}(t), \ldots, \mathbf{a}_{j_{d(i)}}(t)\right]^T \in \mathbb{R}^{md(i)}$ in linear form as

$$\mathbf{a}_{n(i)}(t) = H_{i,t}^T\mathbf{x} = [U_{j_1,t}L_{j_1,t}; \ldots; U_{j_{d(i)},t}L_{j_{d(i)},t}]\mathbf{x} \tag{48}$$

where $H_{i,t}^T = [U_{j_1,t}L_{j_1,t}; \ldots; U_{j_{d(i)},t}L_{j_{d(i)},t}] \in \mathbb{R}^{md(i) \times Nm}$ is the observation matrix of agent i.

Agent i's belief of $\mathbf{x}$ at time t is normal, and at time $t+1$ agent i observes a linear combination of $\mathbf{x}$. Hence, agent i's belief at time $t+1$ can be obtained by a sequential LMMSE update. As a result, mean estimates remain weighted sums of private signals as in (42). In the following lemma, we explicitly present the way we compute the estimation weights, $L_{i,t+1}$ and $Q_{i,t+1}$, at time $t+1$ when $\boldsymbol{\theta} \in \mathbb{R}^m$.

Lemma 4 *Consider a Bayesian game with quadratic function as in* (40) *and the same assumptions and definitions of Lemma 3. Further define the gain matrices as*

$$K_{\mathbf{x}}^i(t) := M_{\mathbf{xx}}^i(t)H_{i,t}\left(H_{i,t}^T M_{\mathbf{xx}}^i(t)H_{i,t}\right)^{-1}, \tag{49}$$

$$K_{\boldsymbol{\theta}}^i(t) := M_{\boldsymbol{\theta}\mathbf{x}}^i(t)H_{i,t}\left(H_{i,t}^T M_{\mathbf{xx}}^i(t)H_{i,t}\right)^{-1}. \tag{50}$$

If agents play according to a linear equilibrium strategy then agent i's posterior $\mathbf{P}_{i,t+1}([\boldsymbol{\theta}^T, \mathbf{x}^T])$ is Gaussian with means that are linear combination of private signals,

$$\mathbf{E}_{i,t+1}[\boldsymbol{\theta}] = Q_{i,t+1}\mathbf{x}, \text{ and } \mathbf{E}_{i,t+1}[\mathbf{x}] = L_{i,t+1}\mathbf{x}, \tag{51}$$

where the estimation matrices are given by

$$L_{i,t+1} = L_{i,t} + K_{\mathbf{x}}^i(t) \left(H_{i,t}^T - H_{i,t}^T L_{i,t} \right), \tag{52}$$

$$Q_{i,t+1} = Q_{i,t} + K_{\boldsymbol{\theta}}^i(t) \left(H_{i,t}^T - H_{i,t}^T, L_{i,t} \right), \tag{53}$$

and the covariance matrices are further given by

$$M_{\mathbf{xx}}^i(t+1) = M_{\mathbf{xx}}^i(t) - K_{\mathbf{x}}^i(t) H_{i,t}^T M_{\mathbf{xx}}^i(t), \tag{54}$$

$$M_{\boldsymbol{\theta\theta}}^i(t+1) = M_{\boldsymbol{\theta\theta}}^i(t) - \left[K_{\boldsymbol{\theta}}^i(t)^T H_{i,t}^T M_{\mathbf{x\theta}}^i(t) \right]^T, \tag{55}$$

$$M_{\boldsymbol{\theta}\mathbf{x}}^i(t+1) = M_{\boldsymbol{\theta}\mathbf{x}}^i(t) - K_{\boldsymbol{\theta}}^i(t) H_{i,t}^T M_{\mathbf{xx}}^i(t). \tag{56}$$

Proof: The proof is identical to the proof of Lemma 2 with the action coefficients $U_{i,t}$ taking the place of $\mathbf{v}_{i,t}$. ∎

Lemma 4 shows that when mean estimates are linear combinations of private signals at time t, they remain that way at time $t + 1$. In the next theorem, we show that assumption in (42) is indeed true for all time by realizing that the estimates at time $t = 0$ are linear combinations of private signals. To simplify presentation of initial conditions, we assume that agent i's private signals are independent, $\mathbf{E}_{i,0}[\mathbf{x}_i[k]\mathbf{x}_i[l]] = 0$ for all $k = 1, \ldots, m$ and $l \neq k$.

Theorem 2 *Given the quadratic utility function in* (40)*, if there exists a linear equilibrium strategy σ_t^* as in* (43) *for $t \in \mathbb{N}$, then the action coefficients $U_{i,t}$ can be computed by solving the system of linear equations in* (44)*, and further, agents' estimates of $\mathbf{x}$ and $\boldsymbol{\theta}$ are linear combinations of private signals as in* (42) *with estimation matrices computed recursively using* (49)-(50) *and* (52)-(56) *with initial values*

$$Q_{i,0} := \begin{pmatrix} \mathbf{e}_i^T & \mathbf{0}_{1 \times N} & \ldots & \mathbf{0}_{1 \times N} \\ \mathbf{0}_{1 \times N} & \mathbf{e}_i^T & \ldots & \mathbf{0}_{1 \times N} \\ \vdots & \ldots & \ddots & \vdots \\ \mathbf{0}_{1 \times N} & \ldots & \mathbf{0}_{1 \times N} & \mathbf{e}_i^T \end{pmatrix} \in \mathbb{R}^{m \times Nm}, \tag{57}$$

$$L_{i,0} := \operatorname{diag}\left(\left[\mathbf{1}\mathbf{e}_i^T, \ldots, \mathbf{1}\mathbf{e}_i^T \right] \right) \in \mathbb{R}^{Nm \times Nm}, \tag{58}$$

where $\mathbf{e}_i \in \mathbb{R}^N$. The initial covariance matrix $M_{\mathbf{xx}}^i(0) \in \mathbb{R}^{Nm \times Nm}$ is a diagonal block matrix with $N \times N$ blocks $((M_{\mathbf{xx}}^i))_{k,k} \in \mathbb{R}^{N \times N}$ for $k = 1, \ldots, m$, initial variance $M_{\boldsymbol{\theta\theta}}^i(0) \in \mathbb{R}^{m \times m}$ and initial

---

Algorithm 1 QNG filter for $\boldsymbol{\theta} \in \mathbb{R}^d$

Initialization: Set posterior distribution on $\boldsymbol{\theta}$ and $\mathbf{x}$

$$\begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{x} \end{bmatrix} \mid h_{i,0} \sim \mathcal{N} \left(\begin{bmatrix} Q_{i,0}\mathbf{x} \\ L_{i,0}\mathbf{x} \end{bmatrix}, \begin{pmatrix} M^i_{\boldsymbol{\theta\theta}}(0), M^i_{\boldsymbol{\theta}\mathbf{x}}(0) \\ M^i_{\mathbf{x}\boldsymbol{\theta}}(0), M^i_{\mathbf{xx}}(0) \end{pmatrix} \right)$$

and $\{L_{j,0}, \mathbf{k}_{j,0}\}_{j \in V}$ according to (57) and (58).

For $t = 0, 1, 2, \ldots$

1) *Equilibrium strategy:* Solve for $\{U_{j,t}\}_{j \in V}$ using the set of equations in (44).

2) *Play and observe:* Take action $a_i(t) = U_{i,t}\mathbf{E}_{i,t}[\mathbf{x}]$ and observe $a_{n(i)}(t)$.

3) *Observation matrix:* Construct $H_{i,t}$ using (48).

4) *Bayesian estimates:* Update $\mathbf{E}_{i,t}[\mathbf{x}]$ and $\mathbf{E}_{i,t}[\theta]$ using (30) and (31), respectively. Update error covariance matrices using (54)–(56).

5) *Estimation weights:* Update $\{L_{j,t}, \mathbf{k}_{j,t}\}_{j \in V}$ using (52)–(53).

---

cross covariance $M^i_{\boldsymbol{\theta}\mathbf{x}}(0) \in \mathbb{R}^{m \times Nm}$ are given by

$$\left((M^i_{\mathbf{xx}}) \right)_{k,k} = \mathrm{diag}(\bar{\mathbf{e}}_i)\mathrm{diag}(\mathbf{C}_{k,k}) + \bar{\mathbf{e}}_i \bar{\mathbf{e}}_i^T C_i[k,k], \tag{59}$$

$$M^i_{\boldsymbol{\theta\theta}}(0) = C_i, \tag{60}$$

$$M^i_{\boldsymbol{\theta}\mathbf{x}}(0) = C_i \begin{pmatrix} \bar{\mathbf{e}}_i^T & \mathbf{0}_{1 \times N} & \ldots & \mathbf{0}_{1 \times N} \\ \mathbf{0}_{1 \times N} & \bar{\mathbf{e}}_i^T & \ldots & \mathbf{0}_{1 \times N} \\ \vdots & \ldots & \ddots & \vdots \\ \mathbf{0}_{1 \times N} & \ldots & \mathbf{0}_{1 \times N} & \bar{\mathbf{e}}_i^T \end{pmatrix} \tag{61}$$

Proof: See Appendix B. ∎

Similar to the scalar case, when network structure and the equilibrium strategy profile are common knowledge, agent i can calculate the weights $\{U_{j,t}\}_{j \in V}$ for all t and update his estimates locally. In Algorithm 1, we provide a sequential local algorithm for agent i to calculate updates for $\boldsymbol{\theta}$ and $\mathbf{x}$ and to act according to equilibrium strategy. The Bayesian rational learning defined here in Algorithm 1 for the vector state case follows the same steps for the scalar case defined in Section IV and by Figs. 1 and 2.

VI. COURNOT COMPETITION

In a Cournot competition model N firms produce a common good that they sell in a market with limitless demand. The cost per production unit c is common for all firms and constant for all times. The selling unit price, however, decreases as the total amount of goods produced by all companies increases. We adopt the specific linear model $p - \sum_{j \in V} a_j$ for the selling unit price, where p is the constant market
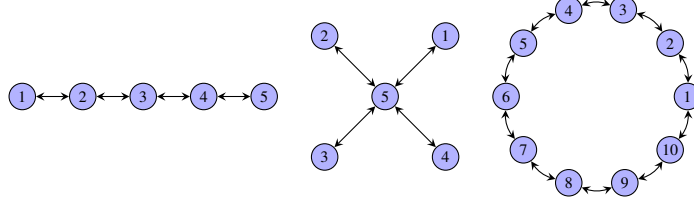
Fig. 3. Line, star and ring networks.

price when no goods are produced. The profit of firm i for production level $a_i \in \mathbb{R}^+$ is therefore given by the utility

$$u_i(a_i, \{a_j\}_{j \in V \setminus i}, \theta) = -ca_i + (p - a_i - \sum_{j \in V \setminus i} a_j)a_i. \tag{62}$$

The utility function in (62) is not of the quadratic form given in (2) because there are two information externalities, the cost c and the clearing price p. While it is possible to resort to the vector form of the QNG filter covered in Section V, it is simpler to write (62) in a form compatible with (2) by defining the parameter $\theta := p - c$ as the effective unit profit at the market price. Using this definition in (62) and reordering terms yields

$$u_i(a_i, \{a_j\}_{j \in V \setminus i}, \theta) = (\theta - a_i - \sum_{j \in V \setminus i} a_j)a_i. \tag{63}$$

Since this utility function is of the form in (2), we can use the QNG filter of Section IV as summarized in Figs. 1 and 2 to determine subsequent BNE production levels. The explicit form of the equilibrium equation in (8) is

$$\sigma_{i,t}^*(h_{i,t}) = \frac{1}{2}\mathbf{E}_{i,t}[\theta] - \frac{1}{2} \sum_{j \in V \setminus i} \mathbf{E}_{i,t}[\sigma_{j,t}^*(h_{j,t})]. \tag{64}$$

It is immediate from (64) that when $\mathbf{E}_{i,t}[\theta] < 0$ it is best for firm i to shut down production. To avoid boundary conditions we restrict attention to cases where private signals $\mathbf{x}$ are such that $\mathbf{E}_{i,t}[\theta] > 0$ for all $i \in V$ and $t \in \mathbb{N}$. This can be guaranteed if all private signals are nonnegative, i.e., $\mathbf{x} \geq \mathbf{0}$. In a game with complete information all private signals $\mathbf{x}$ are known to all agents. In this case the (regular) Nash equilibrium actions of all agents coincide and are given by

$$a_i^* = \frac{\mathbf{E}[\theta \mid \mathbf{x}]}{N + 1} \quad \text{for all } i \in V. \tag{65}$$

The numerical simulations in the next section show that the BNE strategies in (64) converge to the (regular) Nash equilibrium strategy (65) in a finite number of steps.
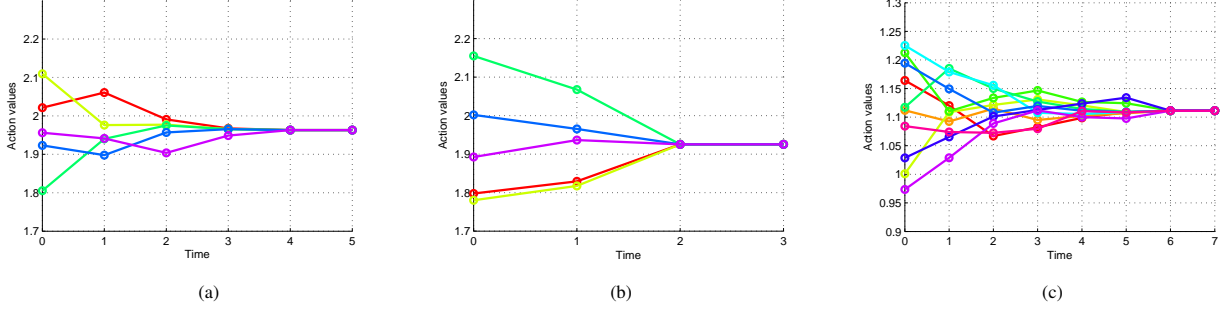
Fig. 4. Agents' actions over time for the Cournot competition game and networks shown in Fig. 3. Each line indicates the quantity produced for an individual at each stage. Actions converge to the Nash equilibrium action of the complete information game in the number of steps equal to the diameter of the network.
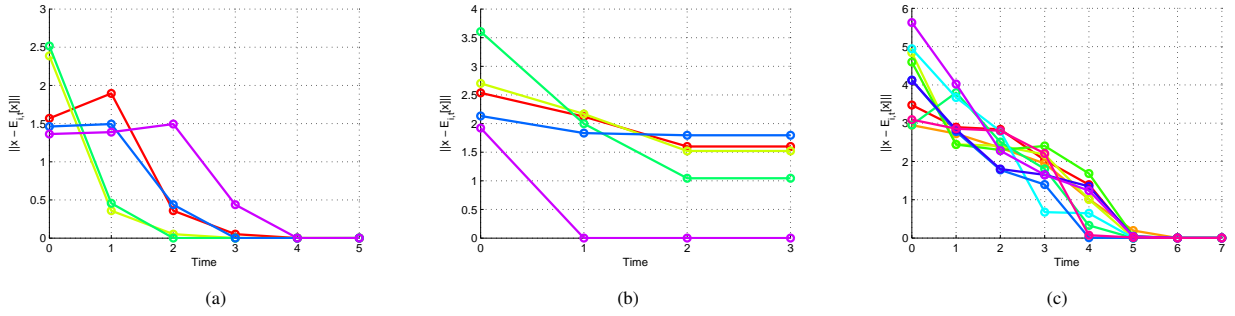

Fig. 5. Normed error in estimates of privates signals, $\|\mathbf{x} - \mathbf{E}_{i,t}[\mathbf{x}]\|_2^2$, for the Cournot competition game and networks shown in Fig. 3. Each line corresponds to an agent's normed error in mean estimates of private signals over the time horizon. While all of the agents learn the true values of all the private signals in line and ring networks, in the star network only the central agent learns all of the private signals.

A. Learning in Cournot competition

The underlying effective unit profit is chosen as $\theta = 12\$/$unit. Firms observe private signals with the additive noise term coming from standard normal distribution, i.e., $\epsilon_i \sim \mathcal{N}(0,1)$. Given this setting, we consider three benchmark networks: a line network with $N = 5$ firms, a star network with $N = 5$ firms, and a ring network with $N = 10$ firms (see Fig. 3).

The quantities produced by firms over time are shown in Fig. 4 for the line (a), star (b) and ring (c) networks. In all of the cases, we observe consensus in the units produced. Furthermore, the consensus production a^* is optimal; that is, firms converge to the Bayes-Nash equilibrium under complete information (65). This implies that all of the firms learn the best estimate of θ by the convergence time T, that is, $\mathbf{E}_{i,T}[\theta \mid h_{i,T}] = \mathbf{E}[\theta \mid \mathbf{x}]$ for all $i \in V$.

Figs. 5(a)–(c) show the error in estimation of private signals $\|\mathbf{x} - \mathbf{E}_{i,t}[\mathbf{x}]\|_2^2$ for all $i \in V$ and $t \in \mathbb{N}$. In Figs. 5(a) and 5(c), corresponding to line and ring networks, the mean square error in private signal estimates goes to zero for all of the firms at the end of the convergence time T. On the other hand, in the
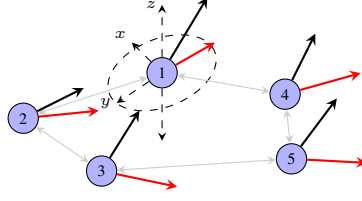
Fig. 6. Mobile agents in a 3-dimensional coordination game. Agents observe initial noisy private signals on heading and take-off angles. Agents revise their estimates on true heading and take-off angles and coordinate their movement angles with each other through local observations.

star network in Fig. 5(b), except for the center firm 5, none of the other firms has zero mean square error in private signal estimates. This means that these firms do not learn at least one of the private signals. As we know from Fig. 4 (b), all of the firms in the star network learn the best estimate of θ given all of the private signals. Hence, in the star network, firms only learn the sufficient statistic to estimate θ (which is the average of the private signals) rather than learning each of the private signals individually.

Figs. 4(a)–(c) suggest that convergence is achieved in $O(\Delta)$ steps where Δ is the diameter of the graph. In [4], it is argued that for the distributed estimation problems when the individual utility function is equal to $u_i(\mathbf{a}_i, \theta) = -(\mathbf{a}_i - \theta)^2$, convergence happens in $O(\Delta)$ steps for tree networks. Our results show that the convergence rate is $O(\Delta)$ not only for tree networks such as line and star networks but also for the ring network when the utility function is quadratic and includes actions of others.

VII. COORDINATION GAME

A network of autonomous agents want to align themselves so that they move toward a goal (x^*, y^*, z^*) on 3-dimensional space following a straight path, and at the same time maintain their initial starting formation. When the goal (x^*, y^*, z^*) is far away, then there exists a common correct direction of movement toward the goal characterized by the heading angle on the $x - y$ plane $\phi \in [0°, 180°]$ and the take-off angle on the $x - z$ plane $\psi \in [0°, 180°]$. Hence, the target movement direction is given by $\boldsymbol{\theta} = [\phi, \psi]^T$. Fig. 6 illustrates a set of autonomous agents on a 3-dimensional plane and their heading and take-off angles where the x, y, z axes are depicted for agent 1.

Mobile agents have the goal of maintaining the starting formation while moving at equal speed by coordinating their movement direction with other agents. Agents need to coordinate with the entire population while communication is restricted to neighboring agents whose direction of movement they can observe. In this context, agent i's decision $\mathbf{a}_i \in [0, 180°] \times [0, 180°]$ represents the heading and take-off angles in the direction of movement. The estimation and coordination goals of agent i can be

represented with the following payoff

$$u_i(\mathbf{a}_i, \{\mathbf{a}_j\}_{j \in V \setminus i}, \boldsymbol{\theta}) = -\frac{1-\lambda}{2}(\mathbf{a}_i - \boldsymbol{\theta})^T(\mathbf{a}_i - \boldsymbol{\theta}) - \frac{\lambda}{2(N-1)}\sum_{j \in V \setminus \{i\}}(\mathbf{a}_i - \mathbf{a}_j)^T(\mathbf{a}_i - \mathbf{a}_j). \qquad (66)$$

The first term is the estimation error in the true heading and take-off angles. The second term is the coordination component that measures the discrepancy between the direction of movement and those of other agents. λ is a constant in $(0, 1)$ gauging the importance of estimation term with respect to the coordination term.

The same payoff formulation can be motivated by looking at learning in organizations [30]. In an organization, individuals share a set of common tasks and have the incentive to coordinate with other units. Each individual receives a private piece of information about the task that needs to be performed while only being able to share his information with whom he has a direct contact in the organization.

Note that the utility function is of the quadratic form given in (40) with vector states and vector actions. Hence, we can use the QNG filter in Section V as summarized in Algorithm 1. As postulated in (8), the explicit equilibrium equation for all $i \in V$ is

$$\sigma_{i,t}^*(h_{i,t}) = (1-\lambda)\mathbf{E}_{i,t}[\boldsymbol{\theta}] + \frac{\lambda}{N-1}\sum_{j \in V \setminus \{i\}}\mathbf{E}_{i,t}[\sigma_{j,t}^*(h_{j,t})]. \qquad (67)$$

In a game with complete information, the Bayes-Nash equilibrium actions of all agents coincide and are given by

$$a_i^* = \mathbf{E}[\boldsymbol{\theta} \mid \mathbf{x}]. \qquad (68)$$

In the next section, we show that the equilibrium actions in (67) converge to the Bayes-Nash equilibrium with complete information as given by (68) in finite number of steps.

A. Learning in coordination games

The correct direction vector is chosen to be $\theta = [10°, 20°]^T$. We let $\lambda = 0.5$. The noise terms, ϵ_i are jointly Gaussian with mean zero and covariance matrix equal to the identity matrix. Having an identity covariance matrix implies that $\mathbf{E}[\mathbf{x}_i[1]\mathbf{x}_i[2]] = 0$.

We evaluate equilibrium behavior in geometric and random networks with $N = 50$ agents, Figs. 7 (a) and (b), respectively. Geometric random network is created by placing the agents randomly on a 4 meter $\times$ 4 meter square and connecting pairs with distance less than 1 meter between them. In the random network, any pair of agents are neighbors with probability 0.1. The geometric network in Fig. 7 (a) has a diameter of $\Delta_g = 5$ where the random network in Fig. 7 (b) has a diameter of $\Delta_r = 4$.
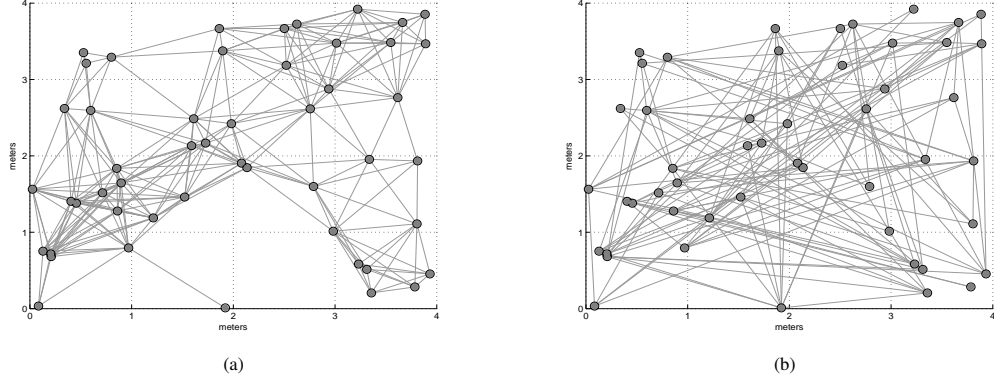
Fig. 7. Geometric (a) and random (b) networks with $N = 50$ agents. Agents are randomly placed on a 4 meter $\times$ 4 meter square. There exists an edge between any pair of agents with distance less than 1 meter apart in the geometric network. In the random network, the connection probability between any pair of agents is independent and equal to 0.1.

The direction of movement of each agent over time is depicted in Figs. 8(a)–(d). Figs. 8(a) and 8(b) show the heading angle ϕ_i of agents in geometric and random networks, respectively. Figs. 8(c) and 8(d) show the take-off angle ψ_i of agents in geometric and random networks, respectively. Fig. 8 illustrates that agents' movement directions converge to the best estimates in heading and take-off angles in a finite number of steps. As a result, at the end of the convergence time T, we have $\mathbf{E}_{i,t}[\phi \,|\, h_{i,T}] = \mathbf{E}[\phi \,|\, \mathbf{x}[1]]$ and $\mathbf{E}_{i,t}[\psi \,|\, h_{i,T}] = \mathbf{E}[\psi \,|\, \mathbf{x}[2]]$ for all $i \in V$. Further, convergence time is in the order of the diameter for both of the networks. This means that agents learn the sufficient statistic to calculate best estimates in the amount of time it takes for information to propagate through the network.

VIII. CONCLUSION

In this paper we introduced the QNG filter that agents can run locally to update their beliefs and select equilibrium actions actions in repeated quadratic games with both information and payoff externalities. The QNG filter provides a mechanism to update beliefs in a Bayes' way when agents' initial prior over the state of the world is Gaussian. We began by showing that when the prior estimates of private signals are Gaussian with means equal to a linear combination of private signals, and the equilibrium strategies of agents are linear combination of mean estimates of private signals, Bayesian updates of estimates of private signals and the underlying state follow a sequential LMMSE estimator. This meant that the estimates remain linear combinations of private signals, and hence, Gaussian. By induction, estimates remain Gaussian for all times if equilibrium actions that are linear in mean of the estimates exist at all the stages. Further, we derived an explicit recursion for tracking of estimates of private signals and calculating equilibrium actions which we leverage to develop the QNG filter. We then extended the QNG
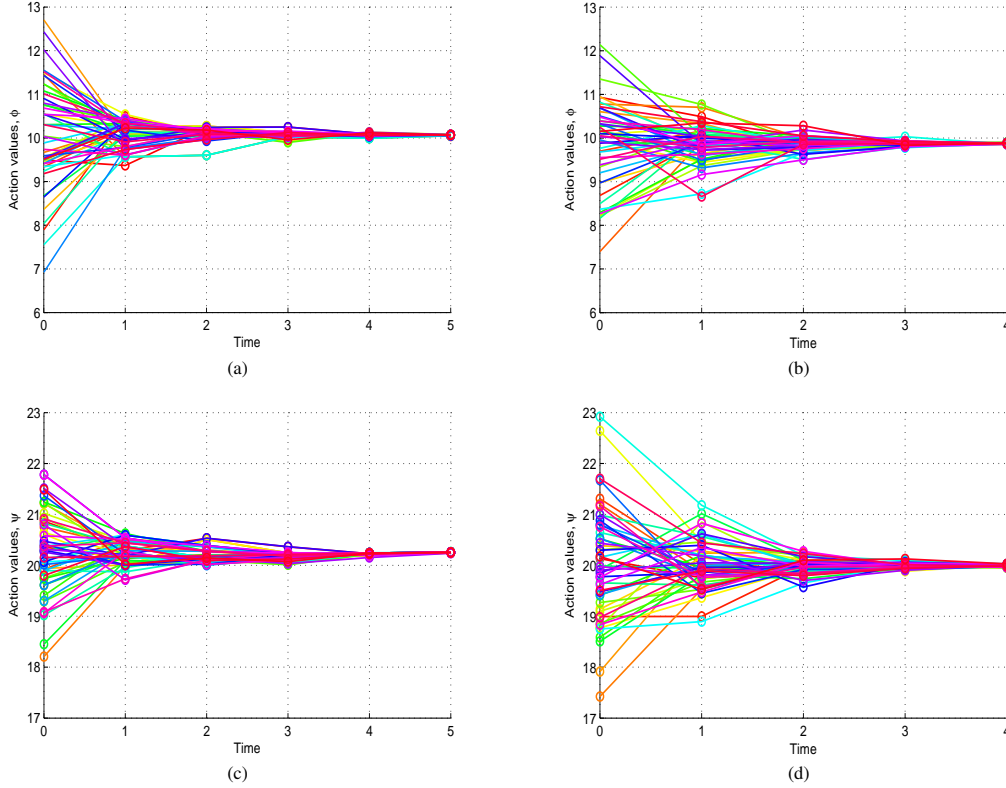
Fig. 8. Agents' actions over time for the coordination game and networks shown in Fig. 7. Values of agents' actions over time for heading angle ϕ_i (top) and take-off angle ψ_i in geometric (left) and random (right) networks respectively. Action consensus happens in the order of the diameter of the corresponding networks.

filter to the case when the state of the world is a vector. We exemplified the QNG filter in Cournot competition game and coordination of mobile agents on 3-dimensional space. In the former the state of the world, effective profit, was a scalar, whereas in the latter the state of the world was a vector including heading and take-off angles. In both examples, the QNG filter converged to the BNE of the game in number of steps that is equal to the order of the diameter of the network. This meant that agents learnt the sufficient statistic of the state while not necessarily learning all the individual private signals.

APPENDIX A

PROOF OF THEOREM 1:

At time $t = 0$ beliefs are normal and have the form in (12). Indeed, since the only information available to agent i at time $t = 0$ is the private signal x_i it follows from the linear observation model in (1) that this is the value assigned to the estimate of all private signals as well as to the estimate of the state θ,

$$\mathbf{E}_{i,0}\left[x_j\right] = x_i \text{ for all } j, \quad \mathbf{E}_{i,0}\left[\theta\right] = x_i. \tag{69}$$

The elements of the matrix $L_{i,0} = \mathbf{1}\mathbf{e}_i^T$ are 1 in the ith column and 0 otherwise. Therefore, the first expression in (69) is equivalent to the first expression in (34). Likewise, since the ith element of $\mathbf{e}_i$ is one with remaining elements zero, the second expression in (69) is equivalent to the second expression in (34). As for the variances in (35), note that the initial estimate of $\mathbf{x}$ has error covariance matrix defined as in (9) for $t = 0$. By substituting initial mean estimates inside (9) and then using the fact that $\mathbf{e}_i^T\mathbf{x} = x_i$, the error covariance matrix can be rewritten as

$$M_{\mathbf{xx}}^i(0) = \mathbf{E}_{i,0}\Big[\big(\mathbf{x} - \mathbf{1}x_i\big)\big(\mathbf{x} - \mathbf{1}x_i\big)^T\Big] \tag{70}$$

From (70), we get the following by using the fact that $x_j - x_i = \epsilon_j - \epsilon_i$ by (1),

$$M_{\mathbf{xx}}^i(0) = \mathbf{E}_{i,0}\Big[\big(\boldsymbol{\epsilon} - \mathbf{1}\epsilon_i\big)\big(\boldsymbol{\epsilon} - \mathbf{1}\epsilon_i\big)^T\Big]. \tag{71}$$

When we expand the terms in (71), we obtain the following

$$M_{\mathbf{xx}}^i(0) = \mathbf{E}_{i,0}\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T\right] - \mathbf{E}_{i,0}\left[\boldsymbol{\epsilon}\mathbf{1}^T\epsilon_i\right] - \mathbf{E}_{i,0}\left[\mathbf{1}\epsilon_i\boldsymbol{\epsilon}^T\right] + \mathbf{1}\mathbf{1}^T\mathbf{E}_{i,0}\left[\epsilon_i^2\right] \tag{72}$$

$$= \mathrm{diag}(\mathbf{c}) - \mathbf{e}_i\mathbf{1}^Tc_i - \mathbf{1}\mathbf{e}_i^Tc_i + \mathbf{1}\mathbf{1}^Tc_i \tag{73}$$

$$= \mathrm{diag}(\mathbf{c}) + \bar{\mathbf{e}}_i\bar{\mathbf{e}}_i^Tc_i - \mathbf{e}_i\mathbf{e}_i^Tc_i \tag{74}$$

Since private signals are independent among agents, that is $\mathbf{E}_{i,0}[\epsilon_k\epsilon_j] = 0$ for all $j \in V \setminus k$ and $k \in V$, we have $\mathbf{E}_{i,0}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T] = \mathrm{diag}(\mathbf{c})$, $\mathbf{E}_{i,0}[\boldsymbol{\epsilon}\epsilon_i] = \mathbf{e}_ic_i$. Using these relations and the definition of noise variance $c_i = \mathbf{E}[\epsilon_i^2]$, (73) follows from (72). When second and third terms are subtracted from the fourth term in (73), we obtain the last two terms in (74). Now, observe that $\mathrm{diag}(\mathbf{c}) - \mathbf{e}_i\mathbf{e}_i^Tc_i = \mathrm{diag}(\bar{\mathbf{e}}_i)\mathrm{diag}(\mathbf{c})$, hence (74) can be rewritten as in (35).

Consider the variance of θ defined in (10) at time $t = 0$. Substituting $\mathbf{E}_{i,0}[\theta] = x_i$ inside (10), we have

$$M_{\theta\theta}^i(0) = \mathbf{E}_{i,0}\left[(\theta - x_i)^2\right] \tag{75}$$

By the signal structure (1) with additive zero mean Gaussian term ϵ_i, we have $\theta - x_i = -\epsilon_i$. As a result, $M_{\theta\theta}^i(0) = \mathbf{E}_{i,0}[\epsilon_i^2]$ which is in return equal to c_i. Next consider the cross-covariance between θ and $\mathbf{x}$ defined in (11) at time $t = 0$,

$$M_{\theta\mathbf{x}}^i(0) = \mathbf{E}_{i,0}\Big[\big(\theta - \mathbf{E}_{i,0}\left[\theta\right]\big)\big(\mathbf{x} - \mathbf{E}_{i,0}\left[\mathbf{x}\right]\big)^T\Big] \tag{76}$$

$$= \mathbf{E}_{i,0}\Big[(-\epsilon_i)(\boldsymbol{\epsilon} - \mathbf{1}\epsilon_i)^T\Big] \tag{77}$$

The second equality follows by substitution of initial mean estimates and then using the definition of private signals (1). Next, we multiply out the terms in (77), use independence of private signals between agents to get (35).

The inductive hypotheses is then true at time $t = 0$ with the explicit initializations in (34) and (35). Lemma 2 has already shown that if the inductive hypothesis is true at time t, it is also true at time $t + 1$. It also provided the explicit recursions in (22)-(23) and (25)-(29). Lemma 1 further shows that the action coefficients $\mathbf{v}_{i,t}$ can be computed by solving the system of linear equations in (15).

APPENDIX B

PROOF OF THEOREM 2:

At time $t = 0$, agents beliefs are normal and have the form in (42). Since the only information available to agent i at time $t = 0$ is the private signal $\mathbf{x}_i$, it follows from the observation model in (38) that agent i assigns $\mathbf{x}_i$ as his mean estimates of the underlying parameter vector and the private signals as in (57)-(58). Next, consider the initial error covariance matrix $M_{\mathbf{xx}}^i(0)$,

$$M_{\mathbf{xx}}^i(0) = \mathbf{E}_{i,0}\left[(\mathbf{x} - \mathbf{E}_{i,0}[\mathbf{x}]) (\mathbf{x} - \mathbf{E}_{i,0}[\mathbf{x}])^T \right] \tag{78}$$

$$= \mathbf{E}_{i,0}\left[\begin{pmatrix} \mathbf{x}[1] - \mathbf{1}\mathbf{x}_i[1] \\ \vdots \\ \mathbf{x}[N] - \mathbf{1}\mathbf{x}_i[N] \end{pmatrix} \begin{pmatrix} \mathbf{x}[1] - \mathbf{1}\mathbf{x}_i[1] \\ \vdots \\ \mathbf{x}[N] - \mathbf{1}\mathbf{x}_i[N] \end{pmatrix}^T \right] \tag{79}$$

Substituting initial mean estimates (58) in (78) and using the fact that $\mathbf{1}\mathbf{e}_i^T \mathbf{x}[n] = \mathbf{1}\mathbf{x}_i[n]$, we get (79). Let $\boldsymbol{\epsilon}[n] := [\epsilon_1[n], \ldots, \epsilon_N[n]]^T \in \mathbb{R}^N$ denote the noise values of agents on the nth state of the world, then we can write each $N \times N$ block of the matrix obtained in (79) as follows

$$\mathbf{E}_{i,0}\left[(\mathbf{x}[k] - \mathbf{1}\mathbf{x}_i[k])(\mathbf{x}[l] - \mathbf{1}\mathbf{x}_i[l])^T \right] = \mathbf{E}_{i,0}\left[(\boldsymbol{\epsilon}[k] - \mathbf{1}\epsilon_i[k]) (\boldsymbol{\epsilon}[l] - \mathbf{1}\epsilon_i[l])^T \right]. \tag{80}$$

Since initial private signals of agent i are assumed to be independent of each other, that is, $\mathbf{E}_{i,0}[\epsilon_i[k]\epsilon_i[l]] = 0$ for all $k = 1, \ldots, m$ and $l \neq k$, (80) is zero when $k \neq l$. When $k = l$, (80) is equivalent to (71). As a result, for the $N \times N$ blocks at the diagonals of $M_{\mathbf{xx}}^i(0)$, we obtain (59) which is similar to its scalar counterpart given in (35). Consider the variance of $\boldsymbol{\theta}$ at time $t = 0$. Using (57), we obtain that $M_{\boldsymbol{\theta\theta}}^i(0)$ is as given in (60). The initial cross covariance can also be calculated using initial mean estimates in (57) and (58) in a similar way.

Given the normal prior $\mathbf{P}_{i,0}([\boldsymbol{\theta}^T, \mathbf{x}^T])$ with mean estimates given by (57)-(58), the inductive hypothesis in Lemma 3 is satisfied at time $t = 0$. Further, by our assumption there exists a linear equilibrium action with weights $U_{i,0}$ that can be calculated by solving the set of equations in (44). Lemma 4 already provides a way to propagate beliefs when agents play according to linear equilibrium strategy. Furthermore, by Lemma 4, if the inductive hypothesis is true at time t then it is also true at time $t + 1$.

REFERENCES

[1] C. Eksin, P. Molavi, A. Ribeiro, and A. Jadbabaie. Bayesian quadratic network game filters. submitted to ICASSP 2013, Vancouver, Canada.

[2] S. Morris and H.S. Shin. The social value of public information. *American Economic Review*, 92:1521–1534, 2002.

[3] J. Sobel. Economists' models of learning. *Journal of Economic Theory*, 94(2):241–261, 2000.

[4] E. Mossel and O. Tamuz. Efficient bayesian learning in social networks with gaussian estimators. *arXiv preprint arXiv:1002.0747*, 2010.

[5] V. Bala and S. Goyal. Learning from neighbours. *Review of Economic Studies*, 65(3), 1998.

[6] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi. Non-Bayesian social learning. *Games and Economic Behavior*, 76(4):210–225, 2012.

[7] P. M. DeMarzo, D. Vayanos, and J. Zwiebel. Persuasion bias, social influence, and unidimensional opinions. *The Quarterly Journal of Economics*, 118:909–968, 2003.

[8] B. Golub and M.O. Jackson. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2:112–149, 2010.

[9] A. Banerjee and D. Fudenberg. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[10] J.J. Xiao, A. Ribeiro, L. Zhi-Quan, and G.B. Giannakis. Distributed compression-estimation using wireless sensor networks. *IEEE Signal Process. Mag.*, 23:27–41, July, 2006.

[11] M. Rabbat, R. Nowak, and J. Bucklew. Generalized consensus computation in networked systems with erasure links. In *Proc. of IEEE 6th Workshop on the Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1088–1092, New York, NY, USA., June 2005.

[12] I. Schizas, A. Ribeiro, and G. Giannakis. Consensus in ad hoc wsns with noisy links - part i: distributed estimation of deterministic signals. *IEEE Trans. Signal Process.*, 56(1):1650–1666, January 2008.

[13] S. Kar and J. M. Moura. Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs. *IEEE J. Sel. Topics Signal Process.*, 5(4):674–690, 2011.

[14] F.S. Cattivelli and A.H. Sayed. Modeling bird flight formations using diffusion adaptation. *IEEE Trans. Signal Process.*, 59(5):2038–2051, 2011.

[15] J. Chen and A.H. Sayed. Diffusion adaptation strategies for distributed optimization and learning over networks. *IEEE Trans. Signal Process.*, 60(8):4289–4305, 2012.

[16] J. Chen, S.-Y. Tu, and A.H. Sayed. Distributed optimization via diffusion adaptation. In *Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, San Juan, Puerto Rico, December 2011.

[17] K.R. Rad and A. Tahbaz-Salehi. Distributed parameter estimation in networks. In *Proc. of the 49th IEEE Conference on Decision and Control*, Atlanta, GA, USA, Dec. 2010.

[18] V. Krishnamurthy. Quickest time herding and detection for optimal social learning. arxiv, abs/1003.4972, 2010, 2011.

[19] A. Ribeiro, I.D. Schizas, S.I. Roumeliotis, and G.B. Giannakis. Kalman filtering in wireless sensor networks. *Control Systems Magazine, IEEE*, 30(2):66–86, 2010.

[20] D. Gale and S. Kariv. Bayesian learning in social networks. *Games Econ. Behav.*, 45:329–346, 2003.

[21] D. Rosenberg, E. Solan, and N. Vieille. Informational externalities and emergence of consensus. *Games Econ. Behav.*, 66:979–994, 2009.

[22] P.M. Djuric and Y. Wang. Distributed bayesian learning in multiagent systems. *IEEE Signal Process. Mag.*, 29:65–76, March, 2012.

[23] Y. Kanoria and O. Tamuz. Tractable bayesian social learning. *Preprint*, 2011.

[24] C. Eksin, P. Molavi, A. Ribeiro, and A. Jadbabaie. Learning in linear games over networks. In *Proceedings of the 50th Annual Allerton Conference on Communications, Control, and Computing*, Allerton, Illinois, USA., 2012.

[25] C. Eksin, P. Molavi, A. Ribeiro, and A. Jadbabaie. Games with side information in economic networks. In *Proceedings of the 46th Asilomar Conference on Signals, Systems and Computers (to appear)*, Pacific Grove, California, USA., 2012.

[26] C. Eksin, P. Molavi, A. Ribeiro, and A. Jadbabaie. Learning in networks with incomplete information. *IEEE Signal Process. Mag.*, (to appear), 2012.

[27] S.M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, Englewood Cliffs, New Jersey, 1. edition, 1993.

[28] T. Ui. Bayesian potentials and information structures: Team decision problems revisited. *International Journal of Economic Theory*, 5(3):271–291, 2009.

[29] R. Radner. Team decision problems. *The Annals of Mathematical Statistics*, 33(3):857–881, 1962.

[30] A. Calvó-Armengol and J.M. Beltran. Information gathering in organizations: equilibrium, welfare, and optimal network structure. *Journal of the European Economic Association*, 7(1):116–161, 2009.